(51) International Patent Classification⁵ : C12N 15/62, 5/00	A1	(11) International Publication Number: WO 91/17253
		(43) International Publication Date: 14 November 1991 (14.11.91)

(21) International Application Number: PCT/DK91/00116

(22) International Filing Date: 30 April 1991 (30.04.91)

(30) Priority data:
1064/90 30 April 1990 (30.04.90) DK

(71) Applicant (for all designated States except US): NOVO NORDISK A/S [DK/DK]; Novo Allé, DK-2880 Bagsværd (DK).

(72) Inventors; and
(75) Inventors/Applicants (for US only) : WIBERG, Finn, Christoph [DK/DK]; Vinkelvej 16, DK-3520 Farum (DK). OLSEN, Frank [DK/DK]; Christoffers Allé 139, DK-2800 Lyngby (DK). BOEL, Esper [DK/DK]; Lyngbakkevej 25, DK-2840 Holte (DK). KJELDSEN, Thomas, Børglum [DK/DK]; Kvædevej 73, DK-2830 Virum (DK). MØLLER, Niels, Peter, Hundahl [DK/DK]; Lyngen 20 A, DK-2800 Lyngby (DK).

(74) Common Representative: NOVO NORDISK A/S; Patent Department, Novo Allé, DK-2880 Bagsvaerd (DK).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.

(54) Title: MAMMALIAN CELLS EXPRESSING A HYBRID RECEPTOR

(57) Abstract

Mammalian cells containing a hybrid DNA insert which comprises a first DNA sequence encoding part of the extracellular domain of a first cellular receptor and a second DNA sequence encoding part of the extracellular domain of second cellular receptor which is specific for a different ligand than the first cellular receptor have favourable growth properties in serum-containing as well as serum-free media.

MAMMALIAN CELLS EXPRESSING A HYBRID RECEPTOR

FIELD OF INVENTION

The present invention relates to a cell expressing a hybrid
between different cellular receptors and a process of producing
5 a desired polypeptide from the cell.

BACKGROUND OF THE INVENTION

Cell culture has been carried out for many years in order to
invenstigate for instance the physiology of mammalian cells and
the relationship between mammalian cells and other organisms
10 such as viruses as well as the impact of a large number of
different compounds on the cells. The equipment tranditionally
used for cell culture includes glass vessels, petri dishes, T-
flasks and the like. The media formulations used for cell
culture experiments have been named for the persons who
15 developed them (often on a trial-and-error basis) or modified
them in various ways, i.a. Ham, Dulbecco, Leibovitz, McCoy,
Waymouth, Eagle, Iscove, etc.

These conventional media all contain serum which was found to
make the cells grow at a reasonable rate for experimentation
20 purposes, and the serum which has generally been found to
provide the best growth rate is foetal calf serum. The scarcity
and variable quality of this serum has created serious problems
for cell culture in that the experiments have often been
difficult to reproduce.

25 Although other aspects of cell culture, in particular the
equipment used to culture the cells on a larger scale, have
undergone a rapid development in recent years, the composition
of the standard media used for cultivation is substantially the
same as in the early days of cell culture. Thus, the media
30 still have a high serum content.

Modern concepts of producing pharmaceuticals on the basis of cell culture require a stringent process control, not only with respect to purification, but also with respect to culture conditions. The use of serum in media is inconsistent with this
5 requirement: due to its biological nature, the composition of serum varies from batch to batch resulting in variability of the cultivation process. Serum also represents the single most important source of unrelated protein present in the culture liquid, which can only be removed from the product of interest
10 by laborious methods resulting in a lower yield of the product of interest. Moreover, serum is the single most important source of contamination of the cell culture with mycoplasma, bacteria or viruses, and costly as well as time-consuming measures have to be taken to remove these contaminants from the
15 serum and to test for their absence.

It would therefore be highly desirable to omit serum from media used for mammalian cell culture. Attempts have been made to adapt cell lines already used in the production of pharmaceuticals or potentially useful cell lines to growth in serum-free
20 media or media with a reduced serum content. It has, however, been observed that when adaptations to serum-free or low serum media have been made, this has been at the expense of growth performance, i.e. the cells grow at a slower rate than in serum-containing media, resulting in lower productivity and
25 higher susceptibility to invading contaminants.

A cell line which is able to grow at an increased growth rate than normally observed in serum-containing media, not only in serum-containing media, but also in serum-free media would therefore be of great potential interest for cell culture in
30 the pharmaceutical industry. By omitting serum, it would be simpler to purify the product of interest resulting in improved yields and purity, and a stable faster cell growth rate would lead to a high process productivity, i.a. because rapid-

growing cells would be less susceptible to infectious organisms.

SUMMARY OF THE INVENTION

It has surprisingly been found that cells with favourable
5 growth properties in serum-containing as well as serum-free media may be obtained by introducing into the cells a DNA construct which comprises a hybrid DNA sequence coding for parts of two different cell surface receptors.

Accordingly, the present invention relates to a mammalian cell
10 containing a hybrid DNA insert which comprises a first DNA sequence encoding part of the extracellular domain of a first cellular receptor and a second DNA sequence encoding part of the extracellular domain of a second cellular receptor which is specific for a different ligand than the first cellular
15 receptor.

Cellular receptors which are typically located on cell surfaces mediate certain important cell to cell interactions involving the transmission of extracellular signals by the interaction with ligands. Such receptors are composed of an
20 extracellular domain (in monomeric or dimeric form) which is capable of specifically recognizing and binding a particular ligand and which may have highly glycosylated and protease-resistant structure, a transmembrane domain which is responsible for anchoring the receptor in the cell membrane and
25 which consists of a hydrophobic sequence of some 25 amino acids, and a cytoplasmic domain which is responsible for generating a cellular signal as a response to the binding of the ligand to the extracellular domain; the cytoplasmic domain(s) may define an enzymatic activity which is triggered
30 by ligand binding.

The reason why hybrid receptors of the invention may promote the growth of cells (particularly in serum-free media) has not

yet been determined. It is, however, currently assumed that such hybrid receptor constructions provide a constant stimulation of the growth regulatory system of the cells so that their dependence on growth factors (e.g. those present in serum) is

5 reduced. Furthermore, it is possible that hybrid receptors are less sensitive to non-activating ligands (e.g. degradation products or metabolites from the cells) and therefore less likely to be inhibited for binding than the natural receptor.

A number of naturally occurring receptors have previously been

10 identified. Thus, Rubin et al., J. Immun. 135, 1985, pp. 3172-3177, describe the release of large quantities of the inter-leukin-2 receptor (IL-2-R) into the culture medium of activated T-cells. The DNA sequence coding for the insulin receptor has been published by A. Ullrich et al., Nature 313, 28 Feb. 1985,

15 pp. 756-761, and S. Seino, Proc. Natl. Acad. Sci. USA 86, 1989, pp. 114-118. Similarly, the DNA sequence coding for the IGF-I receptor has been published by A. Ullrich et al., The EMBO J. 5(10), 1986, pp. 2503-2512.

European Patent Application, Publication No. 244 221 discloses

20 receptors which are hybrids between the extracellular ligand-binding domain of one receptor fused to a heterologous reporter polypeptide which may be the cytoplasmic domain of another receptor.

B. Kobilka et al, Science 240, 3 June 1988, pp. 1310-1316,

25 disclose chimeras between α2 and β2 adrenergic receptors constructed by replacing DNA sequences encoding various parts of the membrane-spanning domain of one of the receptors by the DNA sequences encoding corresponding parts of the other receptor and expressing the sequences in Xenopus laevis

30 oocytes.

I. Lax et al., The EMBO J. 8(2), 1989, pp. 421-427, disclose chimeras between chicken and human EGF receptors constructed by replacing DNA sequences encoding various parts of one of the

receptors by DNA sequences encoding the corresponding parts of the other receptor, and expressed on the surface of mammalian cells.

However, none of these prior publications discloses any use of
5 either natural or hybrid receptors for promoting the growth of mammalian cells.

In the present context, the expression "specific for a different ligand" is intended to indicate that the second receptor specifically binds a ligand which, in a given organism, has a
10 different biological function and/or activity than the ligand binding to the first receptor. The expression is intended to distinguish the present hybrid receptors from those described by, for instance, I. Lax et al., op. cit., which are hybrids of receptors specific for essentially the same ligand derived from
15 two different organisms. The term "ligand" may be defined as a substance which, in nature, is capable of binding to a particular cellular receptor. Preferred ligands are those which act in a similar way as the natural ligand for the receptor in question (e.g. a hormone, growth factor, cytokine or cell
20 adhesion molecule).

DETAILED DISCLOSURE OF THE INVENTION

In particular, the DNA insert introduced into the cell of the present invention is one in which the first DNA sequence (encoding part of the extracellular domain of the first
25 receptor) encodes an exon or a fragment thereof, or in which the second DNA sequence (encoding part of the extracellular domain of the second receptor) encodes an exon or a fragment thereof. For the present purpose, either the first or the second DNA sequence may encode a ligand-binding site of the
30 first or second receptor.

In a particularly preferred embodiment of the cell of the present invention, the DNA insert is one in which the DNA

sequences coding for the first and second receptors exhibit a
high degree of homology in the organization and/or structure of
their exons. This may be advantageous as it is possible to
systematically substitute specific fragments from one receptor
5 for the corresponding fragments from the other receptor. When
producing a cell of the invention, it may therefore be an
advantage to initially replace one or more exons from one of
the receptors by the corresponding exon or exons from the other
receptor, and to test the growth properties of cells into which
10 the DNA construct coding for each hybrid has been introduced.

In this embodiment of the DNA insert introduced in the cell of
the invention, a DNA sequence encoding one or more exons, or a
fragment thereof, of the extracellular domain of the first
receptor may be replaced by a DNA sequence encoding the
15 corresponding exon or exons, or fragment thereof, of the
extracellular domain of the second receptor. Alternatively, a
DNA sequence encoding one or more exons, or a fragment thereof,
of the extracellular domain of the second receptor may be
replaced by a DNA sequence encoding the corresponding exon or
20 exons, or fragment thereof, of the extracellular domain of the
first receptor.

As indicated above, the parent receptors are cell surface
receptors, in particular receptors for hormones, growth
factors, cytokines or cell adhesion molecules. Thus, the
25 receptors may be selected from the group consisting of the
insulin, insulin-like growth factor (IGF), platelet-derived
growth factor (PDGF), epidermal growth factor (EGF), trans-
forming growth factor (TGF), including TGF-α and TGF-ß, growth
hormone and prolactin receptors. The hybrid receptor may
30 advantageously be one comprising at least part of the binding
site for a ligand produced by the cell itself (e.g. containing
the whole or part of the EGF binding site of the EGF receptor).
This permits the cell to stimulate its own growth, thereby
reducing the need for adding external growth promoting agents
35 to the culture medium in order to stimulate cell growth.

In the above-described preferred embodiment of the DNA insert of the invention, hybrids are constructed from receptors which exhibit the aforementioned homology in the overall organization of their DNA.

5 More specifically, in the DNA insert introduced in the cell of the invention, the DNA sequence coding for exon 2, or a fragment thereof, of the insulin receptor may be replaced by the DNA sequence coding for exon 2, or a fragment thereof, of the IGF receptor. Alternatively, the DNA sequence coding for

10 exon 3, or a fragment thereof, of the insulin receptor may be replaced by the DNA sequence coding for exon 3, or a fragment thereof, of the IGF receptor. In a further alternative embodiment, the DNA sequence coding for exons 2 and 3, or a fragment thereof, of the insulin receptor may be replaced by

15 the DNA sequence coding for exons 2 and 3, or a fragment thereof, of the IGF receptor.

In another embodiment of the DNA insert, the DNA sequence coding for exon 2, or a fragment thereof, of the IGF receptor may be replaced by the DNA sequence coding for exon 2, or a

20 fragment thereof, of the insulin receptor. Alternatively, the DNA sequence coding for exon 3, or a fragment thereof, of the IGF receptor may be replaced by the DNA sequence coding for exon 3, or a fragment thereof, of the insulin receptor. As a further alternative, the DNA sequence coding for exons 2 and 3,

25 or a fragment thereof, of the IGF receptor may be replaced by the DNA sequence coding for exons 2 and 3, or a fragment thereof, of the insulin receptor.

Similarly, a DNA sequence coding for exon 1 (preferably combined with exon 2 or a fragment thereof) or one or more of

30 exons 4-11 or fragments thereof of the insulin receptor may be replaced by the corresponding DNA sequence from the IGF receptor, or <u>vice versa</u>.

One example of a cell of the invention exhibiting favourable growth properties is one in which a DNA insert has been introduced, which has the partial DNA sequence shown in Fig. 4A-4F (encoding the extracellular domain of the hybrid recep-
5 tor), or a suitable modification thereof. Suitable modifications of the DNA sequence may comprise nucleotide substitutions which do not give rise to another amino acid sequence of the hybrid polypeptide, but which facilitate the production of the polypeptide, or nucleotide substitutions which do give rise to
10 a different amino acid sequence of the hybrid polypeptide. Other possible modifications may be insertion of one or more nucleotides into the sequence, addition of one or more nucleo-tides at either end of the sequence and deletion of one or more nucleotides at either end of or within the sequence.

15 When the first receptor is the PDGF receptor, the second receptor may suitably be the EGF receptor or vice versa. When the first receptor is the EGF receptor, the second receptor may also be the TGF (α or β) receptor, and vice versa. When the first receptor is the growth hormone receptor, the second
20 receptor may suitably be the prolactin receptor or vice versa.

The DNA insert encoding the hybrid receptor may be prepared synthetically by established standard methods, e.g. the phosphoamidite method described by S.L. Beaucage and M.H. Caruthers, Tetrahedron Letters 22, 1981, pp. 1859-1869, or the
25 method described by Matthes et al., EMBO Journal 3, 1984, pp. 801-805. According to the phosphoamidite method, oligonucleo-tides are synthesized, e.g. in an automatic DNA synthesizer, purified, annealed, ligated and cloned in suitable vectors.

The DNA insert may also be of genomic or cDNA origin, for
30 instance obtained by preparing a genomic or cDNA library and screening for DNA sequences coding for all or part of the hybrid receptor polypeptide by hybridization using synthetic oligonucleotide probes in accordance with standard techniques (cf. Sambrook et al., Molecular Cloning: A Laboratory Manual,

2nd Ed., Cold Spring Harbor, 1989). In this case, a genomic or cDNA sequence encoding a part of either receptor may be modified at a site corresponding to the site(s) at which it is desired to introduce amino acid substitutions, e.g. by site-
5 directed mutagenesis using synthetic oligonucleotides encoding the desired amino acid sequence for homologous recombination in accordance with well-known procedures.

Finally, the DNA insert may be of mixed synthetic and genomic, mixed synthetic and cDNA or mixed genomic and cDNA origin
10 prepared by annealing fragments of synthetic, genomic or cDNA origin (as appropriate), the fragments corresponding to various parts of the entire DNA insert, in accordance with standard techniques. The DNA insert may also be prepared by polymerase chain reaction using specific primers, for instance as de-
15 scribed in US 4,683,202.

The hybrid DNA insert may be introduced into the cell by transfection of the cell with a recombinant expression vector comprising the insert. The expression vector may be any vector which may conveniently be subjected to recombinant DNA pro-
20 cedures, and the choice of vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, i.e. a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid.
25 Alternatively, the vector may be one which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

In the vector, the hybrid DNA insert should be operably
30 connected to a suitable promoter sequence. The promoter may be any DNA sequence which shows transcriptional activity in the host cell of choice and may be derived from genes encoding proteins either homologous or heterologous to the host cell. Examples of suitable promoters for directing the transcription

of the DNA insert in mammalian cells are the SV 40 promoter
(Subramani et al., Mol. Cell Biol. 1, 1981, pp. 854-864), the
MT-1 (metallothionein gene) promoter (Palmiter et al., Science
222, 1983, pp. 809-814) or the adenovirus 2 major late pro-
5 moter.

The DNA insert should also be operably connected to a suitable
terminator, such as the human growth hormone terminator
(Palmiter et al., op. cit.). The vector may further comprise
elements such as polyadenylation signals (e.g. from SV 40 or
10 the adenovirus 5 Elb region), transcriptional enhancer se-
quences (e.g. the SV 40 enhancer) and translational enhancer
sequences (e.g. the ones encoding adenovirus VA RNAs).

The recombinant expression vector may also comprise a DNA
sequence enabling the vector to replicate in the host cell in
15 question. An examples of such a sequence is the SV 40 origin of
replication. The vector may also comprise a selectable marker,
e.g. a gene the product of which complements a defect in the
host cell, such as the gene coding for dihydrofolate reductase
(DHFR) or one which confers resistance to a drug, e.g. neo-
20 mycin, hygromycin or methotrexate.

The procedures used to ligate the hybrid DNA insert with the
DNA sequences coding for the promoter and the terminator,
respectively, and to insert them into suitable vectors con-
taining the information necessary for replication, are well
25 known to persons skilled in the art (cf., for instance,
Sambrook et al., op.cit.).

The cell of the present invention may suitably be used for the
production of proteins of polypeptides of interest. Thus, the
cell may further contain an inserted DNA sequence encoding a
30 desired polypeptide. Examples of such polypeptides are the
human blood clotting factors IX, VIII or VII, human tissue-
type plasminogen activator, human protein C, human plasminogen,
etc.

Examples of suitable mammalian cells for transfection with expression vector containing the DNA insert described above are the COS (ATCC CRL 1650), BHK (ATCC CRL 1632, ATCC CCL 10) NIH/3T3 (ATCC CRL 1658) or CHO (ATCC CCL 61) cell lines.

5 Methods of transfecting mammalian cells and expressing DNA sequences introduced in the cells are described in e.g. Kaufman and Sharp, J. Mol. Biol. 159, 1982, pp. 601-621; Southern and Berg, J. Mol. Appl. Genet. 1, 1982, pp. 327-341; Loyter et al., Proc. Natl. Acad. Sci. USA 79, 1982, pp. 422-426; Wigler et

10 al., Cell 14, 1978, p. 725; Corsaro and Pearson, Somatic Cell Genetics 7, 1981, p. 603, Graham and van der Eb, Virology 52, 1973, p. 456; and Neumann et al., EMBO J. 1, 1982, pp. 841-845.

In another aspect, the present invention relates to a process

15 for producing a desired polypeptide, which comprises culturing a cell as described above in a suitable nutrient medium under conditions which are conducive to the expression of the polypeptide, and recovering the polypeptide from the culture. Although the medium used to culture the cells may be any

20 conventional medium suitable for growing mammalian cells, the cell of the invention has surprisingly been found to exhibit exceptionally favourable growth properties in serum-free medium which, as indicated above, is an advantage, i.a. because the purification of the desired polypeptide will be simplified, and

25 the yield of the polypeptide will be improved.

The polypeptide produced by the cells in this manner will often be secreted to the growth medium and may be recovered from the medium by conventional procedures including separating the host cells from the medium by centrifugation or filtration, preci-

30 pitating the proteinaceous components of the supernatant or filtrate by means of a salt, e.g. ammonium sulphate, followed by purification by a variety of chromatographic procedures, e.g. ion exchange chromatography, affinity chromatography, or the like.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is further described with reference to the drawings, wherein

Fig. 1A-1I shows the full-length cDNA sequence of the human
5 insulin receptor;

Fig. 2 shows the assembly into the α-subunit and part of the β-subunit of the IGF-I receptor (sIGF-I-R) of cDNA fragments generated by reverse transcription and PCR using specific oligonucleotide primers; the circled numerals refer to the
10 primers listed in Example 1;

Fig. 3A-3H shows the cDNA sequence of the soluble IGF-I receptor and the deduced amino acid sequence thereof, indicated in the conventional one-letter code;

Fig. 4A-4I shows the cDNA sequence of the extracellular domain
15 of a hybrid insulin/IGF-I receptor, wherein exons 2 and 3 of the insulin receptor have been replaced by exons 2 and 3 of the IGF-I receptor, and the deduced amino acid sequence thereof, indicated in the conventional one-letter code;

Fig. 5 is a graph showing the growth curves (day 3-7) of three
20 different cell lines, BHK, FCW 38-16 (BHK transfected with DNA encoding the hybrid insulin/IGF-I receptor according to Example 1) and hIR 12-14 (BHK transfected with DNA coding for the wild-type human insulin receptor), in medium containing 5% foetal calf serum;

25 Fig. 6 is a graph showing the growth curves (day 1-3) of three different cell lines, BHK, FCW 38-16 (BHK transfected with DNA encoding the hybrid insulin/IGF-I receptor according to Example 1) and hIR 12-14 (BHK transfected with DNA coding for the wild-

type human insulin receptor), in medium containing 5% foetal calf serum; and

Fig. 7 is a graph showing the growth curves (day 1-8) of three different cell lines, BHK, FCW 38-16 (BHK transfected with DNA encoding the hybrid insulin/IGF-I receptor according to Example 1) and hIR 12-14 (BHK transfected with DNA coding for the wild-type human insulin receptor), in serum-free medium supplemented with human insulin (BHK cells were also grown in serum-free medium without any added human insulin).

The invention is further illustrated in the following examples which are not in any way intended to limit the scope of the present invention.

EXAMPLE 1

Construction of cDNA encoding the human insulin receptor

Insulin receptor cDNA was isolated from a cDNA library generated from poly(A)$^+$ mRNA isolated from the human lympho-blastoid cell line IM9 (available from the American Type Culture Collection, Rockville, Maryland, under the catalogue number ATCC CCL 159) stimulated with 1.4 μg/ml cortisol for 20 hours, substantially according to the method described by Okayama and Berg, <u>Mol. Cell Biol.</u> <u>2</u>, 1982, p. 161 ff.; Okayama and Berg, <u>Mol. Cell. Biol.</u> <u>3</u>(2), Feb. 1983, pp. 280-289; and Noma et al., <u>Nature</u> <u>319</u>, 1986, p. 640 ff. An approximately 4000 bp clone containing the 3'end of the insulin receptor was isolated. To obtain a full-length clone, primer extension was applied on mRNA from the IM9 cells using (10 μg total mRNA) AMV reverse transcriptase (60 U, available from Pharmacia LKB Biotechnology, Sweden) with 800 ng of the 3' oligonucleotide primer 5'-CATCTCAGCAAGATCTTGTCA-3'. The second strand was synthesized as described by Okayama and Berg, <u>op. cit.</u>, using the 5' oligonucleotide 5'-ACCGGGAGCGCGCGCTCTGATC-3' as primer. The cDNA was digested with the restriction endonucleases BssHII

and BglII and ligated into an appropriate vector. A 1550 bp clone containing the 5'end of the insulin receptor was iso- lated. The full-length insulin receptor was assembled in the mammalian expression vector Zem219b (described in DK Patent

5 Application No. 3023/88) digested with NcoI and XbaI by ligating the 5' cDNA clone digested with NcoI and BglII and the 3' cDNA clone digested with SpaI and BglII. The DNA sequence of the full length insulin receptor cDNA is shown in Fig. 1A-1F.

Construction of cDNA encoding a soluble IGF-I receptor

10 A soluble IGF-I receptor cDNA was prepared from mRNA from human term placenta by a method involving polymerase chain reaction (PCR) using specific oligonucleotide primers (cf. R.K. Saiki et al., Science 239, 29 Jan., 1988, pp. 487-491, and US 4,683,202). mRNA was prepared by standard procedures as

15 described in Sambrook et al., op. cit. The mRNA was reverse transcribed into cDNA with AMV reverse transcriptase (Pharmacia LKB Biotechnology, Sweden) using the appropriate specific primers or an oligo-dT-primer at a final concentration of 800 ng/3.5 μg mRNA.

20 IGF-I receptor cDNA fragments were amplified by PCR using the Gene Amp kit (Perkin Elmer Cetus, Norwalk, CT, USA) as re- commended by the manufacturer.

In each PCR, approximately 1μg of the reverse transcribed mRNA was used as template. The primers used for PCR all contained an

25 endonuclease restriction site permitting subcloning and assembly of the soluble IGF-I receptor. the following specific primers were used for PCR

        1.   5'-CCA AAT AGG ATC CAT GAA CTC TGG CTC CGG AGG-3'
        2.   3'-CCC GAA GTA GGC CTT AAG GTC GGT CTC GT-5'
30 3.   5'-TGG GCA GCT GCA GCG CGC CTG-3'
        4.   3'-ATT GGA CCG TGG CCA TGG CCG-5'
        5.   5'-TAA CCT GGC ACC GGT ACC GGC-3'

6.   3'-TCA AAG AGT TGC TTC GAA GAC-5'

7.   5'-AAC ACC ACG GCC GCA GAC ACC-3'

8.   3'-TGT CCT ATA CTT TTG ATT AGA TCT GAC TAG T-5'

The following restriction sites native to the IGF-I receptor
5 were part of the PCR primers: PstI, Asp718, XmaIII and
HindIII. A BamHI site was incorporated in the primer 1 used to
generate the 5' end of the receptor cDNA by introducing a
mismatch in the sequence. The primer 8 used to generate the 3'
end of the soluble IGF-I receptor cDNA was designed by intro-
10 ducing a mismatch in the sequence so as to include a termina-
tion codon in nucleotide position 2842-2844 followed by an XbaI
site. Each PCR reaction cycle comprised denaturation of the
template at 94°C for 1 minute, and annealing of the primers to
the templates for 2 minutes at 50°C, followed by extension of
15 the primers for 3 minutes at 72°C. This cycle was repeated 25
times, resulting in specific IGF-I receptor fragments.

The isolated cDNA fragments were digested with the endo-
nucleases BamHI and XbaI (New England Biolabs, MA, USA) and
subcloned into the PBSII vector (Stratagene, CA, USA) by the
20 method described by Sambrook et al., op. cit. Cells of E. coli
strain MC1061 (T.V. Huynk et al., in DNA Cloning, Vol. 1 (D.M.
Glover, ed.), IRL Press Ltd., Oxford, England, 1983, pp. 56-
110) and SCS-1 (D. Hanahan, J. Mol. Biol. 166, 1983, pp. 557-
580) were made competent according to the method described by
25 D. Hanahan, in DNA Cloning, Vol. 1, supra, pp. 110-135, and
used for transformation with the vectors indicated above. The
soluble IGF-I receptor cDNA was assembled from four subcloned
cDNA fragments as shown in Fig. 2 (BamHI-PstI, PstI-Asp718,
Asp718-PstI, PstI-XbaI) according to the method described by
30 Sambrook et al., op. cit. The cDNA fragment(s) were sequenced
by the enzymatic chain termination method described by F.
Sanger et al., op. cit., using T4 DNA polymerase (Sequenase
Kit, USB, Cleveland, Ohio, USA). The cDNA sequence of the
soluble IGF-I receptor is shown in Fig. 3A-3E. The sequence was

identical to the published sequence (A. Ullrich et al., The EMBO Journal 5(10), 1986, pp. 2503-2512).

Construction of cDNA coding for a hybrid insulin-IGF-I receptor

cDNA coding for a hybrid receptor composed of the human insulin
5 receptor described above wherein the cDNA sequence coding for exons 2 and 3 was replaced by the cDNA sequence coding for exons 2 and 3 from the IGF-I receptor, was prepared by PCR as described above. Primer 2 (shown above), which includes the insulin receptor "compatible" restriction site EcoRI, was used
10 to generate an IGF-I receptor cDNA fragment corresponding to the DNA sequence encoding exon 2 and 3 and the coding part of exon 1 of the insulin receptor which was inserted into the EcoRI site of the insulin receptor cDNA by the method described by Sambrook et al., op. cit.. The sequence of the cDNA fragment
15 was established as described above, and the sequence of the IGF-I receptor cDNA fragment inserted into the insulin receptor was found to be identical with the published sequence (A. Ullrich et al., op. cit.). The cDNA sequence of the hybrid receptor is shown in Fig. 4A-4F.

20 The cDNA fragment encoding the hybrid insulin-IGF-I receptor constructed as described above was inserted in the mammalian expression vector Zem219b (described in DK Patent Application No. 3023/88) which carries a gene conferring methotrexate resistance to the host cell, and cDNA encoding the hybrid
25 receptor was transfected into BHK cells, resulting in the cell line FCW 38-16. Resistant cells were selected with 0.4-2 μM methotrexate.

The hybrid receptor (sIGF-I-R.1-68) comprising an IGF-I receptor in which the 68 N-terminal amino acids of the insulin
30 receptor α-subunit have replaced the equivalent IGF-I receptor segment was constructed by means of restriction enzyme sites similarly positioned in the two receptor cDNAs. The BamHI-XhoI cDNA fragments were exchanged between the insulin receptor and

the IGF-I receptor resulting in sIGF-I-R.1-68. This hybrid cDNA
was inserted into the mammalian expression vector pZem 219b and
transfected into BHK cells, resulting in the cell line FCW 110.

During cultivation in 5% FCS, the cell line FCW 110 formed
5 grape-like clusters of loosely attached cells, forming many
floating cells that stayed alive, indicating that this cell
line may be adapted to grow in suspension with or without
serum.

EXAMPLE 2

10 **Growth performance of cells transfected with the hybrid
insulin/IGF-I receptor**

Materials and methods

Cell lines: BHK (Baby Hamster Kidney cells, Syrian Hamster)
            FCW 38-16, according to Example 1,
15   hIR 12-14, BHK transfected with DNA coding
            for the wild-type human insulin receptor.

Cells were grown in 1 μM MTX (only the transformed cell lines),
DMEM, 10% FCS, 1% PS, 1% Gln.

In a first experiment, 10^4 cells from each cell line were
20 divided among 25 different 5 cm culture dishes and grown in 5
% FCS on day 0. The cells were counted from 4 different plates
every day. Two independent aliquots were sampled from each
plate; where appropriate, floating or loosely attached cells
were also counted in aspirated medium. Cells grown in this
25 growth curve experiment were propagated in standard medium for
growing mammalian cells.

In a second experiment, 10^5 cells from each cell line were
divided among 25 different 5 cm culture dishes on day 0. The
medium was changed on day 1 to 0% FCS, but was supplemented

with 5 mg/1 of human insulin. The cells were counted over the
next days as described above.

Results

a) Growth in medium with 5% FCS

5 The growth curves for the three different cell lines in FCS-
containing medium are shown in Fig. 5. Based on the counts from
the first three days, the cell lines FCW 38-16 and hIR 12-14
show a doubling time of 9 hrs as opposed to 16 hrs for BHK
cells.

10 This experiment was repeated the following week, but this time
with additional cell counts on day 1 and 2. The growth curves
corresponding to the cell counts are shown i Fig. 6.

When figures from the first two days are used to calculate the
generation time, FCW 38-16 and hIR 12-14 double each 14.4 hr
15 and BHK doubles each 20.4 hr. All three cell lines had the same
cell counts on day four, and were not counted further.

In the second week of experiment, all three cell lines had
longer generation times; the ratio between the generation time
for the transformed cell lines and BHK was 1.4 as compared to
20 1.7 for the first week.

b) Growth in serum-free medium supplemented with human insulin

Three growth experiments with all three cell lines in serum-
free medium supplemented with human insulin were performed over
three weeks, and cell counts from these three experiments were
25 averaged to generate the growth curves shown in Fig. 6.

BHK cells grown without serum and without added human insulin
were counted for one week.

FCW 38-16 consistently grew to higher densities under these conditions, but ultimately died on day 8 - 9.

hIR 12-14 formed grape-like clusters of loosely attached cells, forming many floating cells that stayed alive for 2 - 3 days.

5 **Conclusion**

In two independent growth-curve experiments in serum supplemented with 5% FCS, the transformed cell lines showed a shorter generation time than BHK cells. At the end of each experiment all three cell lines showed the same cell density on
10 the plates.

Although the generation times obtained in both weeks of analysis are not the same, it may be concluded that the growth curves support the observed difference in growth rate of FCW 38-16 (and hIR 12-14) compared to BHK.

15 Under serum-free conditions (supplemented with insulin), the FCW 38-16 cell line performed far better than both the BHK and hIR 12-14 cell lines, supporting the utility of cells of the present invention expressing a hybrid insulin/IGF-I receptor.

CLAIMS

1. A mammalian cell containing a hybrid DNA insert which comprises a first DNA sequence encoding part of the extracellular domain of a first cellular receptor and a second DNA
5 sequence encoding part of the extracellular domain of a second cellular receptor which is specific for a different ligand than the first cellular receptor.

2. A cell according to claim 1, wherein the first DNA sequence encodes an exon or a fragment thereof.

10 3. A cell according to claim 1, wherein the second DNA sequence encodes an exon or a fragment thereof.

4. A cell according to any of claims 1-3, wherein either the first or the second DNA sequence encodes a ligand-binding site of the first or second receptor.

15 5. A cell according to claim 1, wherein the DNA sequences coding for the first and second receptors exhibit a high degree of identity/homology in the organization and/or structure of the exons encoded by said sequences.

6. A cell according to any of claims 1-5, wherein a DNA
20 sequence encoding one or more exons, or a fragment thereof, of the extracellular domain of the first receptor is replaced by a DNA sequence encoding the corresponding exon or exons, or fragment thereof, of the extracellular domain of the second receptor.

25 7. A cell according to any of claims 1-5, wherein a DNA sequence encoding one or more exons, or a fragment thereof, of the extracellular domain of the second receptor is replaced by a DNA sequence encoding the corresponding exon or exons, or fragment thereof, of the extracellular domain of the first
30 receptor.

8. A cell according to any of claims 1-7, wherein the first receptor is the insulin receptor.

9. A cell according to any of claims 1-8, wherein the second receptor is the insulin-like growth factor (IGF) receptor.

5 10. A cell according to any of claims 1-7, wherein the first receptor is the platelet-derived growth factor (PDGF) receptor.

11. A cell according to claim 10, wherein the second receptor is the epidermal growth factor (EGF) receptor.

12. A cell according to any of claims 1-7, wherein the first
10 receptor is the EGF receptor.

13. A cell according to claim 12, wherein the second receptor is the transforming growth factor (TGF) receptor.

14. A cell according to any of claims 1-7, wherein the first receptor is the TGF receptor.

15 15. A cell according to claim 14, wherein the second receptor is the EGF receptor.

16. A cell according to any of claims 1-7, wherein the first receptor is a growth hormone receptor.

17. A cell according to claim 16, wherein the second receptor
20 is a prolactin receptor.

18. A cell according to any of claims 1-7, wherein the first receptor is a prolactin receptor.

19. A cell according to claim 18, wherein the second receptor is a growth hormone receptor.

20. A cell according to claim 8, wherein a DNA sequence encoding exon 2, or a fragment thereof, of the insulin receptor is replaced by a DNA sequence encoding exon 2, or a fragment thereof, of the IGF receptor.

5 21. A cell according to claim 8, wherein a DNA sequence encoding exon 3, or a fragment thereof, of the insulin receptor is replaced by a DNA sequence encoding exon 3, or a fragment thereof, of the IGF receptor.

22. A cell according to claim 8, wherein a DNA sequence
10 encoding exons 2 and 3, or a fragment thereof, of the insulin receptor is replaced by a DNA sequence encoding exons 2 and 3, or a fragment thereof, of the IGF receptor.

23. A cell according to claim 9, wherein a DNA sequence encoding exon 2, or a fragment thereof, of the IGF receptor is
15 replaced by a DNA sequence encoding exon 2, or a fragment thereof, of the insulin receptor.

24. A cell according to claim 9, wherein a DNA sequence encoding exon 3, or a fragment thereof, of the IGF receptor is replaced by a DNA sequence encoding exon 3, or a fragment
20 thereof, of the insulin receptor.

25. A cell according to claim 9, wherein a DNA sequence encoding exons 2 and 3, or a fragment thereof, of the IGF receptor is replaced by a DNA sequence encoding exons 2 and 3, or a fragment thereof, of the insulin receptor.

25 26. A cell according to claim 22, wherein the DNA insert has the partial DNA sequence shown in Fig. 4A-4F (encoding the extracellular domain of the hybrid insulin/IGF-I receptor), or a suitable modification thereof.

27. A cell according to any of claims 1-26, which further contains an inserted DNA sequence encoding a desired poly-peptide.

28. A process for producing a desired polypeptide, which
5 comprises culturing a cell according to claim 27 in a suitable nutrient medium under conditions conducive to the expression of the polypeptide and recovering the polypeptide from the culture.

29. A process according to claim 28, wherein the medium is a
10 serum-free medium.

Fig. 1a

```
601  ATCTGTCCGGTACGCGAAGGGCAAGACCAACTGCCCGCCACCGTCATCAACGGGCAG  660
     ----+----+----+----+----+----+----+----+----+----+----+----+
     TAGACAGGCCCATGGCGCTTCCCGTTCTGGTTGACGGGCGGTGGCAGTAGTTGCCCGTC

661  TTTGTCGAACGATGTTGGACTCATAGTCACTGCCAGAAAGTTGCCCGACCATCTGTAAG  720
     ----+----+----+----+----+----+----+----+----+----+----+----+
     AAACAGCTTGCTACAACCTGAGTATCAGTGACGGTCTTTCAAACGGGCTGGTAGACATTC

721  TCACACGGCTGCACCGCCGAAGGCCTCTGTTGCCACAGCGAGTGCCTGGGCAACTGTTCT  780
     ----+----+----+----+----+----+----+----+----+----+----+----+
     AGTGTGCCGACGTGGCGGCTTCCGGAGACAACGGTGTCGCTCACGACCCGTTGACAAGA

781  CAGCCCGACGACGACCCCCACCAAGTGCCTGTGGCCTGCCCGCAACTTCTACCTGACGGCCAGGTGT  840
     ----+----+----+----+----+----+----+----+----+----+----+----+
     GTCGGGCTGCTGGGGTGGTTCACGCACCGGACGGCGTTGAAGATGACCTGCCGTCCACA

841  GTGGAGACCTGCCCGCCCCCTACTACCACTTCCAGGACTGGCGCTGTGTGAACTTCAGC  900
     ----+----+----+----+----+----+----+----+----+----+----+----+
     CACCTCTGGACGGGCGGGGCATGATGGTGAAGGTCCTGACCGGCGACACACTTGAAGTCG

901  TTCTGCCAGGACCTGCACCACAAATGCAAGAACTCGCGGAGGCAGGGCTGCCACCAGTAC  960
     ----+----+----+----+----+----+----+----+----+----+----+----+
     AAGACGGTCCTGGACGTGGTGTTTACGTTCTTGAGCGGCCTCCGTCCCGACGGTGGTCATG

961  GTCATTCACAACAACAAGTGCATCCCTGAGTGTCCCTCCGGGTACACGATGAATTCCAGC  1020
     ----+----+----+----+----+----+----+----+----+----+----+----+
     CAGTAAGTGTTGTTGTTCACGTAGGGACTCACAGGGAGGCCCATGTGCTACTTAAGGTCG

1021 AACTTGCTGTGCACCCCATGCCTGGGCTCCCGTGTCCCAAGTGTGCCACCTCCTAGAAGGC  1080
     ----+----+----+----+----+----+----+----+----+----+----+----+
     TTGAACGACACGTGGGGTACGGACCCGAGGGCACAGGGTTCCACACGGTGGAGGATCTTCCG

1081 GAGAAGACCATCGACTCGGTGACGTCTGCCCAGGAGCTCCGAGGATGCACCGTCATCAAC  1140
     ----+----+----+----+----+----+----+----+----+----+----+----+
     CTCTTCTGGTAGCTGAGCCACTGCAGACGGGCTCCTACGTGGCAGTAGTTG

1141 GGGAGTCTGATCATCAACATTCGAGGAGGCAACAATCTGGCAGCTGAGCTAGAAGCCAAC  1200
     ----+----+----+----+----+----+----+----+----+----+----+----+
     CCCTCAGACTAGTAGTTGTTAAGCTCCTCCGTTGTTAGACCGTCGACTCGATCTTCGGTTG
```

Fig. 1b

```
1201  CTCGGCCTCATTGAAGAAATTTCAGGGTATCTAAAAATCCGCCGATCCTACGCTCTGGTG
      -----+---------+---------+---------+---------+---------+  1260
      GAGCCGGAGTAACTTCTTAAAGTCCCATAGATTTTAGGCGGCTAGGATGCGAGACCAC

1261  TCACTTTCCTTCTTCCGGAAGTTACGTCTGATTCGAGGAGAGACCTTGGAAATTGGGAAC
      -----+---------+---------+---------+---------+---------+  1320
      AGTGAAAGGAAGAGAGGCCTTCAATGCAGACTAAGCTCCTCTCTGGAACCTTTAACCCTTG

1321  TACTCCTTCTATGCCTTGGACAACCAGAACCTAAGGCAGCTCTGGGACTGGAGCAAACAC
      -----+---------+---------+---------+---------+---------+  1380
      ATGAGGAAGATACGGAACCTGTTGGTCTTGGATTCCGTGAGACCCTGACCTCGTTGTG

1381  AACCTCACCATCACTCAGGGGAAACTCTTCTTCCACTATAACCCCAAACTCTGCTTGTCA
      -----+---------+---------+---------+---------+---------+  1440
      TTGGAGTGGTAGTGAGTCCCTTTGAGAAGAAGGTGATATTGGGGTTTGAGACGAACAGT

1441  GAAATCCACAAGAGATGGAAGAAGTTTCAGGAACCAAGGGCGCCAGGAGAGAAACGACATT
      -----+---------+---------+---------+---------+---------+  1500
      CTTTAGGTGTTCTACCTTCTTCAAAGTCCTTGGTTCCCGCGGTCCTCTCTTTGCTGTAA

1501  GCCCTGAAGACCAATGGGGACCAGGCATCCTGTGAAAATGAGTTACTTAAATTTTCTTAC
      -----+---------+---------+---------+---------+---------+  1560
      CGGGACTTCTGGTTACCCCTGGTCCGTAGGACACTTTTACTCAATGAATTAAAAGAATG

1561  ATTCGGACATCTTTGACAAGATCTTGCTGAGATGGAGCCGTACTGGCCCCCCGACTTC
      -----+---------+---------+---------+---------+---------+  1620
      TAAGCCTGTAGAAAACTGTTCTAGAACGACTCTACCCTGGCATGACCGGGGGCTGAAG

1621  CGAGACCTCTTGGGGTTCATGCTGTTCTACAAAGAGGCCCCTTATCAGAATGTGACGGAG
      -----+---------+---------+---------+---------+---------+  1680
      GCTCTGGAGAACCCCAAGTACGACAAGATGTTTCTCCGGGAATAGTCTTACACTGCCTC

1681  TTCGACGGGCAGGATGCATGTGGTTCCAACAGTTGGACGGTGGTAGACATTGACCCACCC
      -----+---------+---------+---------+---------+---------+  1740
      AAGCTGCCCGTCCTACGTACACCAAGGTTGTCAACCTGCCACCATCTGTAACTGGGTGGG

1741  CTGAGGTCCAACGACCCCAAATCACAGAACCACCCAGGGTGGCTGATGCGGGGTCTCAAG
      -----+---------+---------+---------+---------+---------+  1800
      GACTCCAGGTTGCTGGGGTTTAGTGTCTTGGTGGGTCCCACCGACTACGCCCCAGAGTTC
```

Fig. 1c

```
1801  CCCTGGACCCAGTATGCCATCTTTGTGAAGACCCTGGTCACCTTTTTCGGATGAACGCCGG  1860
      -----+---------+---------+---------+---------+---------+
      GGGACCTGGGTCATACGGTAGAAACACTTCTGGGACCAGTGGAAAAGCCTACTTGCGGCC

1861  ACCTATGGGCCAAGAGTGACATCATTTATGTCCAGACAGATGCCACCAACCCCTCTGTG    1920
      -----+---------+---------+---------+---------+---------+
      TGGATACCCCGGTTCTCACTGTAGTAAATACAGGTCTGTCTACGGTGGTTGGGGAGACAC

1921  CCCCTGGATCCAATCTCAGTGTCTAACTCATCATCCCAGATTATTCTGAAGTGGAAACCA  1980
      -----+---------+---------+---------+---------+---------+
      GGGGACCTAGGTTAGAGTCACAGATTGAGTAGGGTCTAATAAGACTTCACCTTTGGT

1981  CCCTCCGACCCCAATGGCAACATCACCCACTACCTGGTTTTCTGGGAGAGGCAGGCGGAA  2040
      -----+---------+---------+---------+---------+---------+
      GGGAGGCTGGGGTTACCGTTGTAGTGGGTGATGGACCAAAAGACCCTCTCCGTCCGCCTT

2041  GACAGTGAGCTGTTCGAGCTGGATTATTGCCTCAAAGGGCTGAAGCTGCCCTCGAGGACC  2100
      -----+---------+---------+---------+---------+---------+
      CTGTCACTCGACAAGCTCGACCTAATAACGGAGTTTCCCGACTTCGACGGGAGCTCCTGG

2101  TGGTCTCCACCATTCGAGTCTGAAGATTCTCAGAAGCACAACCAGAGTGAGTATGAGGAT  2160
      -----+---------+---------+---------+---------+---------+
      ACCAGAGGTGGTAAGCTCAGACTTCTAAGAGTCTTCGTGTTGGTCTCACTCATACTCCTA

2161  TCGGCCGGCGAATGCTGCTCCTGTCCAAAGACAGACTCTCAGATCCTGAAGGAGCTGGAG  2220
      -----+---------+---------+---------+---------+---------+
      AGCCGGCCGCTTACGACGAGGACAGGTTTCTGTCTGAGAGTCTAGGACTTCCTCGACCTC

2221  GAGTCCTCGTTAGGAAGACGTTTGAGGATTACCTGCACAACGTGGTTTCGTCCCCAGG    2280
      -----+---------+---------+---------+---------+---------+
      CTCAGGAGCAAATCCTTCTGCAAACTCCTAATGGACGTGTTGCACCAAAAGCAGGGGTCC

2281  CCATCTCGGAAACGCAGGTCCCTTGGGCGATGTTGGGAATGTGACGGTGGCCGTGCCCACG  2340
      -----+---------+---------+---------+---------+---------+
      GGTAGAGCCTTTGCGTCCAGGGAACCGCCTACACAACCCTTACACTGCCACCGGCACGGGTGC
```

Fig. 1d

```
2341  GTGGCAGCTTTCCCCAACACTTCCTCGACCAGCGTGCCCAGCGAGTCCGGAGGAGCACAGG  2400
      ---------+---------+---------+---------+---------+---------+
      CACCGTCGAAAGGGGTTGTGAAGGAGCTGGTCGCACGGGTCGCTCAGGCCTCCTCGTGTCC

2401  CCTTTTGAGAAGGTGGTGAACAAGGAGTCGCTGGTCATCTCCGGCTTGCGACACTTCACG   2460
      ---------+---------+---------+---------+---------+---------+
      GGAAAACTCTTCCACCACTTGTTCCTCAGCGACCAGTAGAGGCCGAACGCTGTGAAGTGC

2461  GGCTATCGCATCGAGCTGCAGGCTTGCAACCAGGACACCCCTGAGGAACGGTGCAGTGTG   2520
      ---------+---------+---------+---------+---------+---------+
      CCGATAGCGTAGCTCGACGTCCGAACGTTGGTCCTGTGGGGACTCCTTGCCACGTCACAC

2521  GCAGCCTACGTCAGTGCGAGGACCATGCCTGAAGCCAAGGCTGATGACATTGTTGGCCCT   2580
      ---------+---------+---------+---------+---------+---------+
      CGTCGGATGCAGTCACGCTCCTGGTACGGACTTCGGTTCCGACTACTGTAACAACCGGGA

2581  GTGACGCCATGAAATCTTTGAGAACAACGTCGTCCACTTGATGTGGCAGGAGCCGAAGGAG  2640
      ---------+---------+---------+---------+---------+---------+
      CACTGCGTACTTTAGAAACTCTTGTTGCAGCAGGTGAACTACACCGTCCTCGGCTTCCTC

2641  CCCAATGGTCTGATCGTGCTGTATGAAGTGAGTTATCGGCGATATGGTGATGAGGAGCTG   2700
      ---------+---------+---------+---------+---------+---------+
      GGGTTACCAGACTAGCACGACATACTTCACTCAATAGCCGTATACCACTACTCCTCGAC

2701  CATCTCTGCGACACCCGCAAGCACTTCGCTCTGGAACGGGGCTGCAGGCTGCGTGGGGCTG  2760
      ---------+---------+---------+---------+---------+---------+
      GTAGAGACGCTGTGGGCGTTCGTGTGAAGCGAGACCTTGCCCCGACGTCCGACGCACCCGAC

2761  TCACCGGGGAACTACAGCGTGCGAATCCGGGCGCCACCTCCCTTGCGGGCAACGGCTCTTGG  2820
      ---------+---------+---------+---------+---------+---------+
      AGTGGCCCCTTGATGTCGCACGCTTAGGCCCGGTGGAGGGAAACGCCCGGTTGCCGAGAACC
```

Fig. 1e

```
2821  ACGGAACCCACCTATTTCTACGTGACAGACTATTTAGACGTCCGTCAAATATTGCAAAA  2880
      ---------+---------+---------+---------+---------+---------+
      TGCCTTGGGTGGATAAAGATGCACTGTCTGATAAATCTGCAGGGCAGTTTATAACGTTTT

2881  ATTATCATCGGCCCCCTCATCTTTGTCTTTCTTCAGTGTTGTGATTGGAAGTATTTAT  2940
      ---------+---------+---------+---------+---------+---------+
      TAATAGTAGCCGGGGGAGTAGAAACAGAAAGAGAAGTCACAACACTAACCTTCATAAATA

2941  CTATTCCTGAGAAAGAGGCAGCCAGATGGGCCGCTGGGACCGCTTTACGCTTCTTCAAAC  3000
      ---------+---------+---------+---------+---------+---------+
      GATAAGGACTCTTTCTCCGTCGGTCTACCCGGCGACCCTGGCGAAATGCGAAGAAGTTTG

3001  CCTGAGTATCTCAGTGCCAGTGATGTGTTTCCATGCTCTCTGTGTACGTGCCGGACGAGTGG  3060
      ---------+---------+---------+---------+---------+---------+
      GGACTCATAGAGTCACGGTCACTACACAAAGGTACGAGACACATGCACGGCCTGCTCACC

3061  GAGGTGTCTCGAGAGAAGATCACCCTCCTTCGAGAGCTGGGGGCAGGGCTCCTTCGGCATG  3120
      ---------+---------+---------+---------+---------+---------+
      CTCCACAGAGCTCTCTTCTAGTGGGAGGAAGCTCTCGACCCCGTCCCGAGGAAGCCGTAC

3121  GTGTATGAGGGCAATGCCAGGGACATCATCAAGGGTGAGGCAGAGACCCGCGTGGCGGTG  3180
      ---------+---------+---------+---------+---------+---------+
      CACATACTCCCGTTACGGTCCCTGTAGTAGTTCCCACTCCGTCTCTGGGCGCACCGCCAC

3181  AAGACGGTCAACGAGTCAGCCAGTCTCCGAGAGCGGATTGAGTTCCTCAATGAGGCCTCG  3240
      ---------+---------+---------+---------+---------+---------+
      TTCTGCCAGTTGCTCAGTCGGTCAGAGGCTCTCGCCTAACTCAAGGAGTTACTCCGGAGC

3241  GTCATGAAGGGCTTCACCTGCCATCACGTGGTGCGCCTCCTGGGAGTGGTGTCCAAGGGC  3300
      ---------+---------+---------+---------+---------+---------+
      CAGTACTTCCCGAAGTGGACGGTAGTGCACCACGCGGAGGACCCTCACCACAGGTTCCCG

3301  CAGCCCACGCTGGTGGTGATGAGCTGATGGCTCACGGAGCTGAAGAGCTACCTCCGT  3360
      ---------+---------+---------+---------+---------+---------+
      GTCGGGTGCGACCACCACTACCTCGACTACCGAGTGCCTCTGGACTTCTCGATGGAGGCA

3361  TCTCTGCGGCCAGAGGCTGAGAATAATCCTGGCCGCCCTCCCCCTACCCTTCAAGAGATG  3420
      ---------+---------+---------+---------+---------+---------+
      AGAGACGCCGGTCTCCGACTCTTATTAGGACCGGCGGGATGGGAAGTTCTCTAC
```

Fig. 1f

```
3421  ATTCAGATGGCGGCAGAGATTGCTGACGGGATGGCCTACCTGAACGCCAAGAAGTTTGTG
      ----------+---------+---------+---------+---------+---------+  3480
      TAAGTCTACCGCCGTCTCTAACGACTGCCCTACCGGATGGACTTGCGGTTCTTCAAACAC

3481  CATCGGGACCTGGCAGCGAGAAACTGCATGGTGCCCATGATTTACTGTCAAAATTGGA
      ----------+---------+---------+---------+---------+---------+  3540
      GTAGCCCTGGACCGTGCGCTCTTTGACGTACCACGGGTACTAAAATGACAGTTTTAACCT

3541  GACTTTGGAATGACCAGAGACATCTATGAAACGATTACTACCGGAAAGGGGGCAAGGGT
      ----------+---------+---------+---------+---------+---------+  3600
      CTGAAACCTTACTGGTCTCTGTAGATACTTTGCCTAATGATGGCCTTTCCCCGTTCCCA

3601  CTGCTCCCTGTACGGTGGATGGCACCGGAGTCCCTGAAGGATGGGGTCTTCACCACTTCT
      ----------+---------+---------+---------+---------+---------+  3660
      GACGAGGGACAATGCCACCTACCGTGGCCTCAGGGACTTCCTACCCCAGAAGTGGTGAAGA

3661  TCTGACATGTGGTCTCCTTTGGCGTGGTCCTTTGGGAAATCACCAGCTTGGCAGAACAGCCT
      ----------+---------+---------+---------+---------+---------+  3720
      AGACTGTACACCAGGAAACCGCACCAGGAAACCCTTTAGTGGTCGAACCGTCTTGTCGGA

3721  TACCAAGGCCCTGTCTCTAATGAACAGGTGTTGAAATTTGTCATGGATGGAGGGTATCTGGAT
      ----------+---------+---------+---------+---------+---------+  3780
      ATGGTTCCGGACAGATTACTTGTCCACAACTTTAAACAGTAACTACCTCCCATAGACCTA

3781  CAACCCGACAACTGTCCAGAGAGAGTCACTGACCTCATGCCGCATGTGCTGGCAATTCAAC
      ----------+---------+---------+---------+---------+---------+  3840
      GTTGGGCTGTTGACAGGTCTCTCTCAGTGACTGGAGTACGCGGTACACGACCGTTAAGTTG

3841  CCCAACATGAGGCCAACCTTCCTGGAGATTGTCAACCTGCTCAAGGACGACCTGCACCCC
      ----------+---------+---------+---------+---------+---------+  3900
      GGGTTGTACTCCGGTTGGAAGGACCTCTAACAGTTGGACGAGTTCCTGCTGGACGTGGGG

3901  AGCTTTCCAGAGGTGTCGTTCTTCCACAGCGAGGAGAACAAGGCTCCCGAGAGTGAGGAG
      ----------+---------+---------+---------+---------+---------+  3960
      TCGAAAGGTCTCCACAGCAAGAAGGTGTCGCTCCTCTTGTTCCGAGGGCTCTCACTCCTC

3961  CTGGAGATGGAGTTTGAGGACATGGAGAATGTGCCCCTGACCGTTCCTCGCACTGTCAG
      ----------+---------+---------+---------+---------+---------+  4020
      GACCTCTACCTCAAACTCCTGTACCTCTTATACGGGGACCTGGCAAGGAGCGTGACAGTC
```

Fig. 1g

```
4021  AGGGAGGAGGCGGGGGCCGGGATGGAGGGTCCTCGCTGGGTTCAAGCGGAGCTACGAG
      ----+----+----+----+----+----+----+----+----+----+----+----  4080
      TCCCTCCTCCGCCCCCCGGCCCTACCTCCCAGGAGCGACCCAAAGTTCGCCTCGATGCTC

4081  GAACACATCCCTTACACACACATGAACGCGAGGCAAGAAAACGGGCGGATTCTGACCTTG
      ----+----+----+----+----+----+----+----+----+----+----+----  4140
      CTTGTGTAGGGAATGTGTGTACTTGCCTCCGTTCTTTTGCCCGCCTAAGACTGGAAC

4141  CCTCGGTCCAATCCTTCCTAACAGTGCCTACCGTGGCGGGGCGGGCAGGGGTTCCCATT
      ----+----+----+----+----+----+----+----+----+----+----+----  4200
      GGAGCCAGGTTAGGAAGGATTGTCACGGATGGCACCCGCCCCGTCCCCAAGGGTAA

4201  TTCGCTTTCCTCTGGTTTGAAAGCCCTCTGGAAAACTCAGGATTCTCAGACTCTACCATG
      ----+----+----+----+----+----+----+----+----+----+----+----  4260
      AAGCGAAAGGAGACCAAACTTTCGGAGACCTTTGAGTCCTAAGAGTGCTGAGATGGTAC

4261  TCCAATGGAGTTCAGAGATCGTTCCTATACATTTCTGTTCATCTTAAGGTGGACTCGTTT
      ----+----+----+----+----+----+----+----+----+----+----+----  4320
      AGGTTACCTCAAGTCTCTAGCAAGGATATGTAAAGACAAGTAGAATTCCACCTGAGCAAA

4321  GGTTACCAATTTAACTAGTCCTGCAGAGGATTTAACTGTGAACCTGAAGGGCAAGGGGTT
      ----+----+----+----+----+----+----+----+----+----+----+----  4380
      CCAATGGTTAAATTGATCAGGACGTCTCCTAAATTGACACTTGGACCTCCCGTTCCCCAA

4381  TCCACAGTTGCTGCTCCTTTGGGCAACGACGGTTCAAACCAGGATTTGTGTTTTTTC
      ----+----+----+----+----+----+----+----+----+----+----+----  4440
      AGGTGTCAACGACGAGGAGAAACCCGTTGCTGCCAAAGTTGGTCCTAAAACACAAAAAG

4441  GTTCCCCCCACCCGCCCCCAGCAGATGGAAAAGAGCACCTGTTTTACAAATTCTTTT
      ----+----+----+----+----+----+----+----+----+----+----+----  4500
      CAAGGGGGTGGGCGGGGTCGTCTACCTTTCTTTCGTGGACAAAAATGTTTAAGAAAAA

4501  TTTTTTTTTTTTTGCTGGTGTCTGAGCTTCAGTATAAAAGACAAAACTTCCTGTTTG
      ----+----+----+----+----+----+----+----+----+----+----+----  4560
      AAAAAAAAAAAAAACGACCACAGACTCGAAGTCATATTTCTGTTTGAAGGACAAAC

4561  TGGAACAAAAGTTCGAAAGAAAAACAAAAAACAAAACACCCAGCCCTGTTCCAGGAGAAT
      ----+----+----+----+----+----+----+----+----+----+----+----  4620
      ACCTTGTTTCAAGCTTTCTTTTTTGTTTTGTTTTTGTGGGTCGGGACAAGGTCCTCTTA
```

Fig. 1h

```
4621  TTCAAGTTTTACAGGTTGAGCTTCAAGATGGTTTTTTTGGTTTTTTTTTCTCTCATC
      ----+----+----+----+----+----+  4680
      AAGTTCAAAATGTCCAACTCGAAGTTCTACCAAAAAACCAAAAAAAAAAGAGAGTAG

4681  CAGGCTGAAGGATTTTTTTTTCTTTACAAAATGAGTTCCTCAAATTGACCAATAGCTGC
      ----+----+----+----+----+----+  4740
      GTCCGACTTCCTAAAAAAAAGAAATGTTTTACTCAAGGAGTTTAACTGGTTATCGACG

4741  TGCTTTCATATTTTGGATAAGGGTCTGTGGTCCCGGCGTGTGCTCACGTGTGTATGCACG
      ----+----+----+----+----+----+  4800
      ACGAAAGTATAAAACCTATTCCCAGACACCAGGGCCGCACACGAGTGCACACATACGTGC

4801  TGTGTGTGTCCATTAGACACGGCTGACGTGTGTGCAAAGTATCCATGCGGAGTTGATGCT
      ----+----+----+----+----+----+  4860
      ACACACACAGGTAATCTGTGCCGACTGCACACACGTTTCATAGGTACGCCTCAACTACGA

4861  TTGGGAATTGGCTCATGAAGGTTCTTCTTCAAGGGTGCGAGCTCATCCCCCTCTCTCCTTC
      ----+----+----+----+----+----+  4920
      AACCCTTAACCGAGTACTTCCAAGAGAGTTCCCACGCTCGAGTAGGGGAGAGAGGAAG

4921  CTTCTTATTGACTGGGAGACTGTGTCGCTCTCGACAGATTCTTCTTGTGTGTCAGAAGTCTAGCC
      ----+----+----+----+----+----+  4980
      GAAGAATAACTGACCCTCTGACACGAGAGCTGTGTCTAAGAAGAACACAGTCTTCGAGATCGG

4981  TCAGGTTTCTACCCTCCCCTTCACATTGGTGGCCAAGGGAGGAGCATTTCATTTGGAGTGA
      ----+----+----+----+----+----+  5040
      AGTCCAAAGATGGGAGGGAAGTGTAACCACCGGTTCCCCTCCTCGTAAAGTAAACCTCACT

5041  TTATGAATCTTTTCAAGACCAAACCAAGCTAGGAGACATTAAAAAAAAAAAGAAAAAGA
      ----+----+----+----+----+----+  5100
      AATACTTAGAAAAGTTCTGGTTTGGTTGGATCCTGTAATTTTTTTTTTTTTTTTTCTTTTTCT

5101  AAGAAAAAACAAAATGGAAAAAGGAAAAAAAAAGAACTGAGATGACAGAGTTTGAGA
      ----+----+----+----+----+----+  5160
      TTCTTTTTTTTGTTTTTTACCTTTTTTCCTTTTTTTTTTTTTTTTCTTGACTCTGTCTCAAAACTCT

5161  ATATATTTGTACCATATTT
      ----+------  5179
      TATATAAACATGGTATAAA
```

Fig. 1i
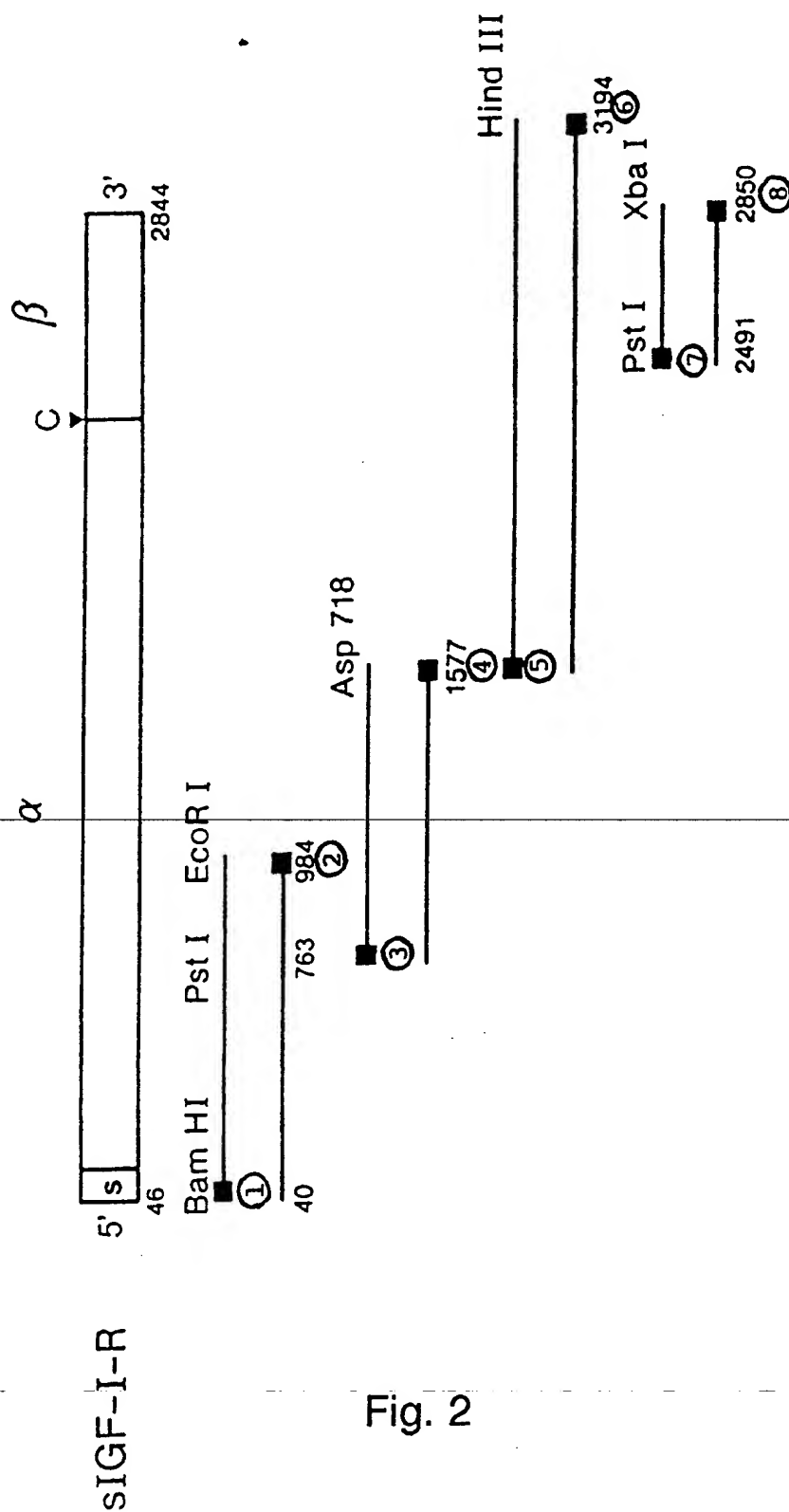
Fig. 2

```
       B
       a
       m
       H
       I
165    GGATCCATGAAGTCTGGCTCCGGAGGAGGGTCCCCGACCTCGCTGTGGGGGCTCCTGTTT   224
       CCTAGGTACTTCAGACCGAGGCCTCCTCCCAGGGGCTGGAGCGACACCCCGAGGACAAA
       G   S   M   K   S   G   S   G   G   G   S   P   T   S   L   W   G   L   L   F

225    CTCTCCGCCGCCTCTCGCTCTGGCCGACGAGTGGAGAAATCTGCGGGCCAGGCATCGAC   284
       GAGAGGCGGCGGAGAGCGAGACCGGCTGCTCACCTCTTTAGACGCCCGGTCCGTAGCTG
       L   S   A   A   L   S   L   W   P   T   S   G   E   I   C   G   P   G   I   D

285    ATCCGCAACGACTATCAGCAGCTGAAGCGCCTGGAGAACTGCACGGTGATCGAGGGCTAC   344
       TAGGCGTTGCTGATAGTCGTCGACTTCGCGGACCTCTTGACGTGCCACTAGCTCCCGATG
       I   R   N   D   Y   Q   Q   L   K   R   L   E   N   C   T   V   I   E   G   Y

345    CTCCACATCCTGCTCATCTCCAAGGCCGAGGACTACCGCAGCTACCGCTTCCCCAAGCTC   404
       GAGGTGTAGGACGAGTAGAGGTTCCGGCTCCTGATGGCGTCGATGGCGAAGGGGTTCGAG
       L   H   I   L   L   I   S   K   A   E   D   Y   R   S   Y   R   F   P   K   L

405    ACGGTCATTACCGAGTACTTGCTGTCGTTCCGAGTGGCCGGCCTCGAGAGCCTCGGAGAC   464
       TGCCAGTAATGGCTCATGAACGACAGCAAGGCTCACCGGCCGGAGCTCTCGGAGCCTCTG
       T   V   I   T   E   Y   L   L   F   R   V   A   G   L   E   S   L   G   D

465    CTCTTCCCCAACCTCACGGTCATCCGCGGCTGGAAACTCTTCTACAACTACGCCCTGGTC   524
       GAGAAGGGGTTGGAGTGCCAGTAGCGCCGACCTTTGAGAAGATGTTGATGCGGGACCAG
       L   F   P   N   L   T   V   I   R   G   W   K   L   F   Y   N   Y   A   L   V
```

Fig. 3a

```
525                                                          584
     ATCTTCGAGATGACCAATCTCAAGGATATTGGGCTTTACAACCTGAGGAACATTACTCGG
     ----+----+----+----+----+----+----+----+----+----+----+----+  -
     TAGAAGCTCTACTGGTTAGAGTTCCTATAACCCGAAATGTTGGACTCCTTGTAATGAGCC
      I  F  E  M  T  N  L  K  D  I  G  L  Y  N  L  R  N  I  T  R

585                                                          644
     GGGGCCATCAGGATTGAGAAAAATGCTGACCTCTGTTACCTCTCCACTGTGGACTGGTCC
     ----+----+----+----+----+----+----+----+----+----+----+----+  -
     CCCCGGTAGTCCTAACTCTTTTACGACTGGAGACAATGGAGAGGTGACACCTGACCAGG
      G  A  I  R  I  E  K  N  A  D  L  C  Y  L  S  T  V  D  W  S

645                                                          704
     CTGATCCTCGGATGCGGTGTCCAATAACTACATTGTGGGAATAAGCCCCAAAGGAATGT
     ----+----+----+----+----+----+----+----+----+----+----+----+  -
     GACTAGGAGCCTACGCCACAGGTTATTGATGTAACACCCCTTATTCGGGGGTTTCCTTACA
      L  I  L  D  A  V  S  N  N  Y  I  V  G  N  K  P  P  K  E  C

705                                                          764
     GGGGACCTGTGTCCAGGGACCATGGAGGAGAAGCCGATGTGTGAGAAGACCACCATCAAC
     ----+----+----+----+----+----+----+----+----+----+----+----+  -
     CCCCTGGACACAGGTCCCTGGTACCTCCTCTTCGGCTACACACTCTTCTGGTGGTAGTTG
      G  D  L  C  P  G  T  M  E  E  K  P  M  C  E  K  T  T  I  N

765                                                          824
     AATGAGTACAACTACCGCTGCTGGACCAACAAACCGCTGCCAGAAAATGTGCCCAAGCACG
     ----+----+----+----+----+----+----+----+----+----+----+----+  -
     TTACTCATGTTGATGGCGACGACCTGGTTGTTTGGCGACGGTCTTTTACACGGGTTCGTGC
      N  E  Y  N  Y  R  C  W  T  N  R  C  Q  K  M  C  P  S  T

825                                                          884
     TGTGGGAAGCGGGCGTGCACCGAGAACAATGAGTGCTGCCACCCCGAGTGCCTGGGCAGC
     ----+----+----+----+----+----+----+----+----+----+----+----+  -
     ACACCCTTCGCCCGCACGTGGCTCTTGTTACTCACGACGGTGGGGCTCACGACCCGTCG
      C  G  K  R  A  C  T  E  N  N  E  C  C  H  P  E  C  L  G  S
```

Fig. 3b

```
885  TGCAGCGCGCCTGACAACGACACGGCCTGTGTAGCTTGCCGCCACTACTACTATGCCGGT  944
     ---------+---------+---------+---------+---------+---------+
     ACGTCGCGCGGACTGTTGCTGTGCCGGACACATCGAACGGCGGGTGATGATGATACGGCCA
      C  S  A  P  D  N  D  T  A  C  V  A  C  R  H  Y  Y  Y  A  G  -

945  GTCTGTGTGCCTGCCCGCCCAACACCTACAGGTTTGAGGGCTGGCGCTGTGTGGAC     1004
     ---------+---------+---------+---------+---------+-------
     CAGACACACGGACGGGCGGGTTGTGGATGTCCAAACTCCCGACCGCGACACACCTG
      V  C  V  P  A  C  P  P  N  T  Y  R  F  E  G  W  R  C  V  D  -

1005 CGTGACTTCTGCGCCAACATCCTCAGCGCCGAGAGCAGCGACTCCGAGGGTTTGTGATC   1064
     ---------+---------+---------+---------+---------+---------+
     GCACTGAAGACGCGGTTGTAGGAGTCGCGGCTCTCGTCGCTGAGGCTCCCAAACACTAG
      R  D  F  C  A  N  I  L  S  A  E  S  S  D  S  E  G  F  V  I  -

1065 CACGACGGGCGAGTGCATGCAGGAGTGCCCCCTCGGGCTTCATCCGCAACGGCAGCCAGAGC  1124
     ---------+---------+---------+---------+---------+---------+
     GTGCTGCCGCTCACGTACGTCCTCACGGGGGAGCCCGAAGTAGGCGTTGCCGTCGGTCTCG
      H  D  G  E  C  M  Q  E  C  P  S  G  F  I  R  N  G  S  Q  S  -

1125 ATGTACTGCATCCCTTGTGAAGGTCCTTGCCCGAAGGTCTGTGAGGAAGAAAAGAAAACA  1184
     ---------+---------+---------+---------+---------+---------+
     TACATGACGTAGGGAACACTTCCAGGAACGGGCTTCCAGACACTCCTTCTTTTCTTTTGT
      M  Y  C  I  P  C  E  G  P  C  P  K  V  C  E  E  E  K  K  T  -

1185 AAGACCATTGATTCTGTTACTTCTGCTCAGATGCTCCAAGGATGCCACCATCTTCAAGGGC  1244
     ---------+---------+---------+---------+---------+---------+
     TTCTGGTAACTAAGACAATGAAGACGAGTCTACGAGGTTCCTACGTGGTAGAAGTTCCCG
      K  T  I  D  S  V  T  S  A  Q  M  L  Q  G  C  T  I  F  K  G  -

1245 AATTTGCTCATTAACATCCGACGGGGGAATAACATTGCTTCAGAGCTGGAGAACTTCATG  1304
     ---------+---------+---------+---------+---------+---------+
     TTAAACGAGTAATTGTAGGCTGCCCCTTATTGTAACGAAGTCTCGACCTCTTGAAGTAC
      N  L  L  I  N  I  R  R  G  N  N  I  A  S  E  L  E  N  F  M  -
```

Fig. 3c

```
1305  GGGCTCATCGAGGTGGTGACGGGCTACGTGAAGATCCGCCATTCTCATGCCTTGGTCTCC
      ---------+---------+---------+---------+---------+---------+   1364
      CCCGAGTAGCTCCACCACTGCCCGATGCACTTCTAGGCGGTAAGAGTACGGAACCAGAGG

       G  L  I  E  V  V  T  G  Y  V  K  I  R  H  S  H  A  L  V  S  -

1365  TTGTCCTTCCTAAAAAACCTTCGCCTCATCCTAGGAGAGGAGCAGCTAGAAGGGAATTAC
      ---------+---------+---------+---------+---------+---------+   1424
      AACAGGAAGGATTTTTTGGAAGCGGAGTAGGATCCTCTCCTCGTCGATCTTCCCTTAATG

       L  S  F  L  K  N  L  R  L  I  L  G  E  E  Q  L  E  G  N  Y  -

1425  TCCTTCTACGTCCTCGACAACCAGAACTTGCAGCAACTGTGGGACTGGGACCACCGCAAC
      ---------+---------+---------+---------+---------+---------+   1484
      AGGAAGATGCAGGAGCTGTTGGTCTTGAACGTCGTTGACACCCTGACCCTGGTGGCGTTG

       S  F  Y  V  L  D  N  Q  N  L  Q  Q  L  W  D  W  D  H  R  N  -

1485  CTGACCATCAAAGCAGGGAAAAATGTACTTTGCTTTCAATCCCAAATTATGTGTTCCGAA
      ---------+---------+---------+---------+---------+---------+   1544
      GACTGGTAGTTTCGTCCCTTTTACATGAAACGAAAGTTAGGTTTAATACACAAAGGCTT

       L  T  I  K  A  G  K  M  Y  F  A  F  N  P  K  L  C  V  S  E  -

1545  ATTTACCGCATGGAGGAAGTGACGGGGACTAAAGGGCGCCAAAGCAAAGGGACATAAAC
      ---------+---------+---------+---------+---------+---------+   1604
      TAAATGGCGTACCTCCTTCACTGCCCCTGATTCCCGCGGTTTCGTTCCCGTATTTG

       I  Y  R  M  E  E  V  T  G  T  K  G  R  Q  S  K  G  D  I  N  -

1605  ACCAGGAACAACGGGAGAGAGCCCTCCTGTGAAGTGACGTCCTGCATTCACCTCCACC
      ---------+---------+---------+---------+---------+---------+   1664
      TGGTCCTTGTTGTGCCCCTCTCTCGGAGGACACTTTCACTGCAGGACGTAAAGTGGAGGTGG

       T  R  N  G  E  R  A  S  C  E  S  D  V  L  H  F  T  S  T  -

1665  ACCACGTCGAAGAATCGCATCATCATAACCTGGCACCGGTACCGGTACCGGCCCCTGACTACAGG
      ---------+---------+---------+---------+---------+---------+   1724
      TGGTGCAGCTTCTTAGCGTAGTAGTATTGGACCGTGGCCATGGCGGGGACTGATGTCC

       T  T  S  K  N  R  I  I  I  T  W  H  R  Y  R  P  P  D  Y  R  -
```

Fig. 3d

```
1725  GATCTCATCAGCTTCACCGTTTACTACAAGGAAGCACCCTTTAAGAATGTCACAGAGTAT
      ----+----+----+----+----+----+----+----+----+----+----+----+ 1784
      CTAGAGTAGTCGAAGTGGCAAATGATGTTCCTTCGTGGGAAATTCTTACAGTGTCTCATA
       D   L   I   S   F   T   V   Y   Y   K   E   A   P   F   K   N   V   T   E   Y   -

1785  GATGGGCAGGATGCCTGCGGCTCCAACAGCTGGAACATGGTGGACGTGGACCTCCCGCCC
      ----+----+----+----+----+----+----+----+----+----+----+----+ 1844
      CTACCCGTCCTACGGACGCCGAGGTTGTCGACCTTGTACCACCTGCACCTGGAGGGCGGG
       D   G   Q   D   A   C   G   S   N   S   W   N   M   V   D   V   D   L   P   P   -

1845  AACAAGGACGTGGAGCCCGGCATCTTACTACATGGGCTGAAGCCCTGGACTCAGTACGCC
      ----+----+----+----+----+----+----+----+----+----+----+----+ 1904
      TTGTTCCTGCACCTCGGGCCGTAGAATGATGTACCCGACTTCGGGACCTGAGTCATGCGG
       N   K   D   V   E   P   G   I   L   L   H   G   L   K   P   W   T   Q   Y   A   -

1905  GTTTACGTCAAGGCTGTGACCCTCACCATGGTGGAGAACGACCATATCCGTGGGCCAAG
      ----+----+----+----+----+----+----+----+----+----+----+----+ 1964
      CAAATGCAGTTCCGACACTGGGAGTGGTACCACCTCTTGCTGGTATAGGCACCCGGTTC
       V   Y   V   K   A   V   T   L   T   M   V   E   N   D   H   I   R   G   A   K   -

1965  AGTGAGATCTTGTACATTCGCACCAATGCTTCAGTTCCTTCCATTCCCTTGGACGTTCTT
      ----+----+----+----+----+----+----+----+----+----+----+----+ 2024
      TCACTCTAGAACATGTAAGCGTGGTTACGAAGTCAAGGAAGGTAAGGGAACCTGCAAGAA
       S   E   I   L   Y   I   R   T   N   A   S   V   P   S   I   P   L   D   V   L   -

2025  TCAGCATCGAACTCCTCTTCTCAGTTAATCGTGAAGTGGAACCCTCCCTCTCTGCCCAAC
      ----+----+----+----+----+----+----+----+----+----+----+----+ 2084
      AGTCGTAGCTTGAGGAGAAGAGTCAATTAGCACTTCACCTTGGGAGGGAGAGACGGGTTG
       S   A   S   N   S   S   S   Q   L   I   V   K   W   N   P   P   S   L   P   N   -
```

Fig. 3e

```
2085 GGCAACCTGAGTTACTACATTGTGCGCTGGCAGCGGCAGCCTCAGGACGGCTACCTTTAC
     ----+----+----+----+----+----+----+----+----+----+----+----+ 2144
     CCGTTGGACTCAATGATGTAACACGCGACCGTCGCCGTCGGAGTCCTGCCGATGGAAATG
      G  N  L  S  Y  Y  I  V  R  W  Q  R  Q  P  Q  D  G  Y  L  Y  -

2145 CGGCACAATTACTGCTCCAAAGACAAAATCCCCATCAGGAAGTATGCCGACGGCACCATC
     ----+----+----+----+----+----+----+----+----+----+----+----+ 2204
     GCCGTGTTAATGACGAGGTTTCTGTTTTAGGGGTAGTCCTTCATACGGCTGCCGTGGTAG
      R  H  N  Y  C  S  K  D  K  I  P  I  R  K  Y  A  D  G  T  I  -

2205 GACATTGAGGAGGTCACAGAGAACCCCAAGACTGAGGTGTGGTGGGGAGAAAGGGCCT
     ----+----+----+----+----+----+----+----+----+----+----+----+ 2264
     CTGTAACTCCTCCAGTGTCTCTTGGGGTTCTGACTCCACACACCCCTCTTTCCCGGA
      D  I  E  E  V  T  E  N  P  K  T  E  V  C  G  G  E  K  G  P  -

2265 TGCTGCGCCCTGCCCCAAAACTGAAGCCGAGAAGCAGGCCGAGAAGGAGGCTGAATAC
     ----+----+----+----+----+----+----+----+----+----+----+----+ 2324
     ACGACGCGGGACGGGGTTTTGACTTCGGCTCTTCGTCCGGCTCTTCCTCCTCCGACTTATG
      C  C  A  C  P  K  T  E  A  E  K  Q  A  E  K  E  E  A  E  Y  -

2325 CGCAAAGTCTTTGAGAATTTCCTGCACAACTCCATCTTCGTGCCCAGACCTGAAAGGAAG
     ----+----+----+----+----+----+----+----+----+----+----+----+ 2384
     GCGTTTCAGAAACTCTTAAAGGACGTGTTGAGGTAGAAGCACGGGTCTGGACTTTCCTTC
      R  K  V  F  E  N  F  L  H  N  S  I  F  V  P  R  P  E  R  K  -

2385 CGGAGAGATGTCATGCAAGTGGCCAACACCACCATGTCCAGCCGAAGCAGGAACACCACG
     ----+----+----+----+----+----+----+----+----+----+----+----+ 2444
     GCCTCTCTACAGTACGTTCACCGGTTGGTGGTGGTACAGGTCGGCTTCGTCCTTGTGGTGC
      R  R  D  V  M  Q  V  A  N  T  T  M  S  S  R  S  R  N  T  T  -
```

Fig. 3f

```
2445  GCCGCAGACACCTACAACATCACCGACCCGGAAGAGCTGGAGACAGAGTACCCTTTCTTT
      ---------+---------+---------+---------+---------+---------+  2504
      CGGCGTCTGTGGATGTTGTAGTGGCTGGGCCTTCTCGACCTCTGTCTCATGGGAAAGAAA

       A  D  T  Y  N  I  T  D  P  E  E  L  E  T  E  Y  P  F  F   -

2505  GAGAGCAGAGTGGATAACAAGGAGAGAACTGTCATTTCTAACCTTCGGCCTTTCACATTG
      ---------+---------+---------+---------+---------+---------+  2564
      CTCTCGTCTCACCTATTGTTCCTCTCTTGACAGTAAAGATTGGAAGCCGGAAAGTGTAAC

       E  S  R  V  D  N  K  E  R  T  V  I  S  N  L  R  P  F  T  L   -

2565  TACCGCATCGATATCCACAGCTGCAACCACGAGGCTGAGAAGCTGGGCTGCAGCGCCTCC
      ---------+---------+---------+---------+---------+---------+  2624
      ATGGCGTAGCTATAGGTGTCGACGTTGGTGCTCCGACTCTTCGACCCGACGTCGCGGAGG

       Y  R  I  D  I  H  S  C  N  H  E  A  E  K  L  G  C  S  A  S   -

2625  AACTTCGTCTTTGCAAGGACTATGCCCGCAGAAGGAGCACAGATGACATTCCTGGGCCAGTG
      ---------+---------+---------+---------+---------+---------+  2684
      TTGAAGCAGAAACGTTCCTGATACGGGCGTCTTCCTCGTGTCTACTGTAAGGACCCGGTCAC

       N  F  V  F  A  R  T  M  P  A  E  G  A  D  D  I  P  G  P  V   -

2685  ACCTGGGAGCCCAAGGCCTGAAAACTCCATCTTTTAAAGTGGCCGGAACCTGAGAATCCC
      ---------+---------+---------+---------+---------+---------+  2744
      TGGACCCCTCGGTTCCGGACTTTTGAGGTAGAAAATTTCACCGGCCTTGGACTCTTAGGG

       T  W  E  P  R  P  E  N  S  I  F  L  K  W  P  E  P  E  N  P   -

2745  AATGGGATTGATTCTAATGTATGAAATAAAATACGGATCACACAAGTTGAGGATCAGCCGAGAA
      ---------+---------+---------+---------+---------+---------+  2804
      TTACCTAACTAAGATTACATACTTTATTTTATGCCTAGTGTTCAACTCCTAGTCGCTCTT

       N  G  L  I  L  M  Y  E  I  K  Y  G  S  Q  V  E  D  Q  R  E   -
```

Fig. 3g

```
2805  TGTGTGTCCAGACAGGAATACAGGAAGTATGGAGGGCCAAGCTAAACCCG  2864
      ----+----+----+----+----+----+----+----+----+----+----+----+
      ACACACAGGTCTGTCCTTATGTCCTTCATACCTCCCCGGTTCGATTGGCCGATTTGGGC
       C  V  S  R  Q  E  Y  R  K  Y  G  G  A  K  L  N  R  L  N  P  -

2865  GGGAACTACACAGCCCGGATTCAGGCCACATCTCTCTCTGGGAATGGGTCGTGTGGACAGAT  2924
      ----+----+----+----+----+----+----+----+----+----+----+----+
      CCCTTGATGTGTCGGGCCTAAGTCCGGTGTAGAGAGACCCTTACCCAGCACCTGTCTA
       G  N  Y  T  A  R  I  Q  A  T  S  L  S  G  N  G  S  W  T  D  -
                                                      X
                                                      b
                                                      a
                                                      I

2925  CCTGTGTTCTTCTATGTCCAGGCCAAAACAGGATATGAAAACTAATCTAGAGGATC  2980
      ----+----+----+----+----+----+----+----+----+----+----+
      GGACACAAGAAGATACAGGTCCGGTTTTGTCCTATACTTTGATTAGATCTCCTAG
       P  V  F  F  Y  V  Q  A  K  T  G  Y  E  N  *
```

Fig. 3h

```
B
a
m
H
I

160
ACCCGGGATCCATGAAGTCTGGCTCCGGAGGAGGTCCCCGACCTTCGGCTGTGGGGCTCC  219
+---------+---------+---------+---------+---------+---------+
TGGGCCCTAGGTACTTCAGACCGAGGCCTCCTCCCAGGGGCTGGAGCGACACCCCCGAGG
      M  K  S  G  G  G  G  S  P  T  S  L  W  G  L  L  -

220
TGTTTCTCTCCGCCGCTCTCGCTCTGGCCGACGAGTGGAGAAATCTGCGGCCAGGCA  279
+---------+---------+---------+---------+---------+---------+
ACAAAGAGAGGCGGCGAGAGCGAGACCGGCTGCTCACCTCTTAGACGCCCGGTCCGT
      F  L  S  A  A  L  S  L  W  P  T  S  G  E  I  C  G  P  G  I  -

280
TCGACATCCGCAACGACTATCAGCAGCTGAAGCGCCTGGAGAACTGCACGGTGATCGAGG  339
+---------+---------+---------+---------+---------+---------+
AGCTGTAGGCGTTGCTGATAGTCGTCGACTTCGCGGACCTCTTGACGTGCCACTAGCTCC
      D  I  R  N  D  Y  Q  Q  L  K  R  L  E  N  C  T  V  I  E  G  -

340
GCTACCTCCACATCCTGCTCATCTCCAAGGCCGAGGACTACCGCAGCTACCGCTTCCCA  399
+---------+---------+---------+---------+---------+---------+
CGATGGAGGTGTAGGACGAGTAGAGGTTCCGGCTCCTGATGGCGTCGATGGCGAAGGGGT
      Y  L  H  I  L  L  I  S  K  A  E  D  Y  R  S  Y  R  F  P  K  -

400
AGCTCACGGTCATTACCGAGTACTTGCTGTTCCGAGTGGCCTGCCTCGAGAGCCTCG  459
+---------+---------+---------+---------+---------+---------+
TCGAGTGCCAGTAATGGCTCATGAACGACGACAAGGCTCACCGACGGAGCTCTCGGAGC
      L  T  V  I  T  E  Y  L  L  F  R  V  A  G  L  E  S  L  G  -

460
GAGACCTCTTCCCCAACCTCACGGTCATCCGCGGCTGGAAACTCTTCTACAACTACGCCC  519
+---------+---------+---------+---------+---------+---------+
CTCTGGAGAAGGGGTTGGAGTGCCAGTAGGCGCCGACCTTTGAGAAGATGTTGATGCGGG
      D  L  F  P  N  L  T  V  I  R  G  W  K  L  F  Y  N  Y  A  L  -
```

Fig. 4a

```
520                                                                579
TGGTCATCTTCGAGATGACCAATCTCAAGGATATTGGGCTTTACAACCTGAGGAACATTA
+--------+--------+--------+--------+--------+
ACCAGTAGAAGCTCTACTGGTTAGAGTTCCTATAACCCGAAATGTTGGACTCCTTGTAAT
     V  I  F  E  M  T  N  L  K  D  I  G  L  Y  N  L  R  N  I  T -

580                                                                639
CTCGGGGGGCCATCAGGATTGAGAAAAATGCTGACCTCTGTTACCTCTCCACTGTGGACT
+--------+--------+--------+--------+--------+
GAGCCCCCCGGTAGTCCTAACTCTTTTACGACTGGAGACAATGGAGAGGTGACACCTGA
     R  G  A  I  R  I  E  K  N  A  D  L  C  Y  L  S  T  V  D  W -

640                                                                699
GGTCCCCTGATCCTGATGCGGCGGTGTCCAATAACTACATTGTGGGAATAAGCCCCAAAGG
+--------+--------+--------+--------+--------+
CCAGGGACTAGGACCTACGCCACAGTTATTGATGTAACACCCCTTATTCGGGGGTTTCC
     S  L  I  L  D  A  V  S  N  N  Y  I  V  G  N  K  P  P  K  E -

700                                                                759
AATGTGGGACCTGTGTCCAGGGACCATGGAGGAGAAGCCCGATGTGTGAGAAGACCACCA
+--------+--------+--------+--------+--------+
TTACACCCCTGGACACAGGTCCCTGGTACCTCCTTCGGCTACACACTCTTTCTGGTGGT
     C  G  D  L  C  P  G  T  M  E  E  K  P  M  C  E  K  T  T  I -

760                                                                819
TCAACAATGAGTACAACTACCGCCTGCTGGACCACCACAAACCCGCTGCCAGAAATGTGCCCAA
+--------+--------+--------+--------+--------+
AGTTGTTACTCATGTTGATGGCGACGACCTGGTGTTTGGCGACGGTCTTTTACACGGGTT
     N  N  E  Y  N  Y  R  C  W  T  T  N  R  C  Q  K  M  C  P  S -

820                                                                879
GCACGTGTGGGAAGCGCGGGCGTGCACCGAGAACAATGAGTGCTGCCACCCCGAGTGCCTGG
+--------+--------+--------+--------+--------+
CGTGCACACCCTTCGCCCGCACGTGGCTCTTGTTACTCACGACGGTGGGGCTCACGGACC
     T  C  G  K  R  A  C  T  E  N  N  E  C  C  H  P  E  C  L  G -
```

Fig. 4b

```
880 GCAGCTGCAGCGGCGCCTGACAACGACACGGCCTGTGTAGCTTGCCGCCACTACTACTATG 939
    +---------+---------+---------+---------+---------+---------+
    CGTCGACGTCGCCGCGGACTGTTGCTGTGCCGGACACATCGAACGGCGGTGATGATGATAC
     S  C  S  A  P  D  N  D  T  A  C  V  A  C  R  H  Y  Y  Y  A  -

940 CCGGTGTCTGTGTGCCTGCCTGCCCCGCCCAACACCTACAGGTTTGAGGGCTGGCGCTGTG 999
    +---------+---------+---------+---------+---------+---------+
    GGCCACAGACACACGGACGGACGGGGCGGGTTGTGGATGTCCAAACTCCCGACCGCGACAC
     G  V  C  V  P  A  C  P  P  N  T  Y  R  F  E  G  W  R  C  V  -

1000 TGGACCGGTGACTTCTGCGCCAACATCCTCAGCGCCGAGAGCAGCGACTCCGAGGGGTTTG 1059
     +---------+---------+---------+---------+---------+---------+
     ACCTGGCACTGAAGACGCGGTTGTAGGAGTCGCGGCTCTCGTCGCTGAGGCTCCCCAAAC
      D  R  D  F  C  A  N  I  L  S  A  E  S  S  D  S  E  G  F  V  -

1060 TGATCCACGACGGCGAGTGCATGCAGGAGTGCCCCTCGGGCTTCATCCGGAATTCCAGCA 1119
     +---------+---------+---------+---------+---------+---------+
     ACTAGGTGCTGCCGCTCACGTACGTCCTCACGGGGAGCCCGAAGTAGGCCTTAAGGTCGT
      I  H  D  G  E  C  M  Q  E  C  P  S  G  F  I  R  N  S  S  N  -

1120 ACTTGCTGTGCACCCCATGCCTGGGTCCCTGTCCCAAGTGTGCCACCTCCTAGAAGGCG 1179
     +---------+---------+---------+---------+---------+---------+
     TGAACGACACGTGGGGTACGGACCCAGGGACAGGGTTCCACACGGTGGAGGATCTTCCGC
      L  L  C  T  P  C  L  G  P  C  P  K  V  C  H  L  L  E  G  E  -

1180 AGAAGACCATCGACTCGGTGACGTCTGCCCAGGAGCTCCGAGGATGCACCGTCATCAACG 1239
     +---------+---------+---------+---------+---------+---------+
     TCTTCTGGTAGCTGAGCCACTGCAGACGGGTCCTCGAGGCTCCTACGTGGCAGTAGTTGC
      K  T  I  D  S  V  T  S  A  Q  E  L  R  G  C  T  V  I  N  G  -
```

Fig. 4c

```
1240  GGAGTCTGATCATCAACATTCGAGGAGGCAACAATCTGGCAGCTGAGCTAGAAGCCAACC  1299
      +---------+---------+---------+---------+---------+---------+
      CCTCAGACTAGTAGTTGTAAGCTCCTCCGTTGTTAGACCGTCGACTCGATCTTCGGTTGG
       S  L  I  I  N  I  R  G  G  N  L  A  A  E  L  E  A  N  L  -

1300  TCGGCCTCATTGAAGAAATTTCAGGGTATCTAAAAAATCCGCCGATCCTACGCTCTGGTGT  1359
      +---------+---------+---------+---------+---------+---------+
      AGCCGGAGTAACTTCTTTAAAGTCCCATAGATTTTTTAGGCGGCTAGGATGCGAGACCACA
       G  L  I  E  E  I  S  G  Y  L  K  I  R  R  S  Y  A  L  V  S  -

1360  CACTTTCCTTCTTCCGGAAGTTACGTCTGATTCGAGGAGAGACCTTGGAAATTGGGAACT  1419
      +---------+---------+---------+---------+---------+---------+
      GTGAAAGGAAGAAGGCCTTCAATGCAGACTAAGCTCCTCTCTGGAACCTTTAACCCTTGA
       L  S  F  F  R  K  L  R  L  I  R  G  E  T  L  E  I  G  N  Y  -

1420  ACTCCTTCTATGCCTTGGACAACCAGAACCTAAGGCAGCTCTGGGACTGGAGCAAACACA  1479
      +---------+---------+---------+---------+---------+---------+
      TGAGGAAGATACGGAACCTGTTGGTCTTGGATTCCGTCGAGACCCTGACCTCGTTTGTGT
       S  F  Y  A  L  D  N  Q  N  L  R  Q  L  W  D  W  S  K  H  N  -

1480  ACCTCACCACCACTCAGGGGAAACTCTTCTTCCACTATAACCCCAAACTCTGCTTGTCAG  1539
      +---------+---------+---------+---------+---------+---------+
      TGGAGTGGTGGTGAGTCCCCTTTGAGAAGAAGGTGATATTGGGGTTTGAGACGAACAGTC
       L  T  T  Q  G  K  L  F  F  H  Y  N  P  K  L  C  L  S  E  -

1540  AAATCCACAAGATGGAAGAAGTTTCAGGAACCAAGGGGCGCCAGGAGAGAAACGACATTG  1599
      +---------+---------+---------+---------+---------+---------+
      TTTAGGTGTTCTACCTTCTTCAAAGTCCTTGGTTCCCCGCGGTCCTCTCTTTGCTGTAAC
       I  H  K  M  E  E  V  S  G  T  K  G  R  Q  E  R  N  D  I  A  -
```

Fig. 4d

```
1600
CCCTGAAGACCAATGGGGACAAGGCATCCTGTGAAAATGAGTTACTTAAATTTCTTACA  1659
+----------+---------+---------+---------+---------+---------+
GGGACTTCTGGTTACCCCTGTTCCGTAGGACACTTTTACTCAATGAATTTAAAAGAATGT
 L  K  T  N  G  D  K  A  S  C  E  N  E  L  L  K  F  S  Y  I  -

1660
TTCGGACATCTTTGACAAGATCTTGCTGAGATGGGAGCCGTACTGGCCCCCCGACTTCC  1719
+----------+---------+---------+---------+---------+---------+
AAGCCTGTAGAAAACTGTTCTAGAACGACTCTACCCTCGGCATGACCGGGGGGCTGAAGG
 R  T  S  F  D  K  I  L  L  R  W  E  P  Y  W  P  P  D  F  R  -

1720
GAGACCTCTTGGGGTTCATGCTGTTCTACAAAGAGGCCCCTTATCAGAATGTGACGGAGT  1779
+----------+---------+---------+---------+---------+---------+
CTCTGGAGAACCCCAAGTACGACAAGATGTTTCTCCGGGAATAGTCTTACACTGCCTCA
 D  L  L  G  F  M  L  F  Y  K  E  A  P  Y  Q  N  V  T  E  F  -

1780
TCGATGGGCAGGATGCCGTGTGGTTCCAACAGTTGGACGGTGGTAGACATTGACCCCACCCC  1839
+----------+---------+---------+---------+---------+---------+
AGCTACCCGTCCTACGGCACACCAAGGTTGTCAACCTGCCACCATCTGTAACTGGGTGGGG
 D  G  Q  D  A  C  G  S  N  S  W  T  V  V  D  I  D  P  P  L  -

1840
TGAGGTCCAACGACCCCAAATCACAGAACCACCCAGGGTGGCTGATGCGGGGTCTCAAGC  1899
+----------+---------+---------+---------+---------+---------+
ACTCCAGGTTGCTGGGGTTTAGTGTCTTGGTGGGTCCCACCGACTACGCCCCAGAGTTCG
 R  S  N  D  P  K  S  Q  N  H  P  G  W  L  M  R  G  L  K  P  -

1900
CCTGGACCCCAGTATGCCATCTTTGTGAAGACCCTGGTCACCTTTTCGGATGAACGCCCGA  1959
+----------+---------+---------+---------+---------+---------+
GGACCTGGGTCATACGGTAGAAACACTTCTGGGACCAGTGGAAAAGCCTACTTGCGGCCT
 W  T  Q  Y  A  I  F  V  K  T  L  V  T  F  S  D  E  R  R  T  -
```

Fig. 4e

```
        CCTATGGGGCCAAGAGTGACATCATTTATGTCCAGACAGATGCCACCAACCCCTCTGTGC 2019
1960  +----+----+----+----+----+----+----+----+----+----+----+
        GGATACCCCGGTTCTCACTGTAGTAAATACAGGTCTGTCTACGGTGGTTGGGGAGACACG
         Y  G  A  K  S  D  I  I  Y  V  Q  T  D  A  T  N  P  S  V  P  -
                         B
                         a
                         m
                         H
                         I

        CCCTGGATCCAATCTCAGTGTCTAACTCATCATCCCAGATTATTCTGAAGTGGAAACCAC 2079
2020  +----+----+----+----+----+----+----+----+----+----+----+
        GGGACCTAGGTTAGAGTCACAGATTGAGTAGTAGGGTCTAATAAGACTTCACCTTTGGTG
         L  D  P  I  S  V  S  N  S  S  Q  I  I  L  K  W  K  P  P  -

        CCTCCGACCCCAATGGCAACATCACCCACTACCTGGTTTCTGGGAGAGGCAGGCGGAAG 2139
2080  +----+----+----+----+----+----+----+----+----+----+----+
        GGAGGCTGGGGTTACCGTTGTAGTGGGTGATGGACCAAAAGACCCTCTCCGTCCGCCTTC
         S  D  P  N  G  N  I  T  H  Y  L  V  F  W  E  R  Q  A  E  D  -

        ACAGTGAGCTGTTCGAGCTGGATTATTGCCTCAAAGGGCTGAAGCTGCCCTCGAGGACCT 2199
2140  +----+----+----+----+----+----+----+----+----+----+----+
        TGTCACTCGACAAGCTCGACCTAATAACGGAGTTTCCCGACTTCGACGGGAGCTCCTGGA
         S  E  L  F  E  L  D  Y  C  L  K  G  L  K  L  P  S  R  T  W  -

        GGTCTCCACCATTCGAGTCTGAAGATTCTCAGAAGCACAACCAGAGTGAGTATGAGGATT 2259
2200  +----+----+----+----+----+----+----+----+----+----+----+
        CCAGAGGTGGTAAGCTCAGACTTCTAAGAGTCTTCGTGTTGGTCTCACTCATACTCCTAA
         S  P  F  F  E  S  E  D  S  Q  K  H  N  Q  S  E  Y  E  D  S  -

        CGGCCGGCGAATGCTGCTCCTGTCCAAAGACAGACTCTCAGATCCTGAAGGAGCTGGAGG 2319
2260  +----+----+----+----+----+----+----+----+----+----+----+
        GCCGGCCGCTTACGACGAGGACAGGTTTCTGTCTGAGAGTCTAGGACTTCCTCGACCTCC
         A  G  E  C  C  S  C  P  K  T  D  S  Q  I  L  K  E  L  E  E  -
```

Fig. 4f

```
2320  AGTCCTCGTTTAGGAAGACGTTTGAGGATTACCTGCACAACGTGGTTTTCGTCCCCAGAA  2379
      +---------+---------+---------+---------+---------+---------+
      TCAGGAGCAAATCCTTCTGCAAACTCCTAATGACGTGTTGCACCAAAAGCAGGGGTCTT
       S  S  F  R  K  T  F  E  D  Y  L  H  N  V  V  F  V  P  R  K  -

2380  AAACCTCTTCAGGCACTGGTGCCGAGGACCCTAGGCCATCTCGGAAACGCAGGTCCCTTG  2439
      +---------+---------+---------+---------+---------+---------+
      TTTGGAGAAGTCCGTGACCACGGCTCCTGGGATCCGGTAGAGCCTTTGCGTCCAGGGAAC
       T  S  S  G  T  G  A  E  D  P  R  P  S  R  K  R  R  S  L  G  -

2440  GCGATGTGTGGGAATGTGACGGTGGCCGTGCCCACGGTGGCAGCTTTCCCCAACACTTCCT  2499
      +---------+---------+---------+---------+---------+---------+
      CGCTACACAACCCTTACACTGCCACCGGCACGGGTGCCACCGTCGAAAGGGGTTGTGAAGGA
       D  V  G  N  V  T  V  A  V  P  T  V  A  A  F  P  N  T  S  S  -

2500  CGACCAGCGTGCCCACGAGTCCGGAGGAGCACAGGCCTTTGAGAAGGTGGTGAACAAGG  2559
      +---------+---------+---------+---------+---------+---------+
      GCTGGTCGCACGGGTGCTCAGGCCTCCTCGTGTCCGGAAAACTCTTCCACCACTTGTTCC
       T  S  V  P  T  S  P  E  E  H  R  P  F  E  E  K  V  V  N  K  E  -

2560  AGTCGCTGGTCATCTCCGGCTTGCGACACTTCACGGGCTATCGAGCTGCAGGCTT  2619
      +---------+---------+---------+---------+---------+---------+
      TCAGCGACCAGTAGAGGCCGAACGCTGTGAAGTGCCCGATAGCTAGTCTCGACGTCCGAA
       S  L  V  I  S  G  L  R  H  F  T  G  Y  R  I  E  L  Q  A  C  -

2620  GCAACCAGGACACCCCTGAGGAACGGTGCAGTGTGGCAGCCTACGTCAGTGCGAGGACCA  2679
      +---------+---------+---------+---------+---------+---------+
      CGTTGGTCCTGTGGGGACTCCTTGCCACGTCACACCGTCGGATGCAGTCACGCTCCTGGT
       N  Q  D  T  P  E  E  R  C  S  V  A  A  Y  V  S  A  R  T  M  -
```

Fig. 4g

```
2680  TGCCTGAAGCCAAGGCTGATGACATTGTTGGCCCTGTGACGCATGAAATCTTTGAGAACA
      +---------+---------+---------+---------+---------+---------+  2739
      ACGGACTTCGGTTCCGACTACTGTAACAACCGGACACTGCGTACTTTAGAAACTCTTGT
       P   E   A   K   A   D   D   I   V   G   P   V   T   H   E   I   F   E   N   N   -

2740  ACGTCGTCCACTTGATGTGGCAGGAGCCCGAAGGAGCCCAATGGTCTGATCGTGCTGTATG
      +---------+---------+---------+---------+---------+---------+  2799
      TGCAGCAGGTGAACTACACCGTCCTCGGCTTCCTCGGGTTACCAGACTAGCACGACATAC
       V   V   H   L   M   W   Q   E   P   K   E   P   N   G   L   I   V   L   Y   E   -

2800  AAGTGAGTTATCGGCGATATGGTGATGAGGAGCTGATGAGGAGCTGCGTCTCCGCAAGCACT
      +---------+---------+---------+---------+---------+---------+  2859
      TTCACTCAATAGCCGCTATACCACTACTCCTCGACGTAGAGAGCGCAGAGGCGTTCGTGA
       V   S   Y   R   R   Y   G   D   E   E   L   H   L   C   V   S   R   K   H   F   -

2860  TCGCTCTGAACGGGGCTGCAGGCTGCGTGGCGTGTCACCGGGGAACTACAGCGTGCGAA
      +---------+---------+---------+---------+---------+---------+  2919
      AGCCGAGACCTTGCCCCGACGTCCGACGCACCCGACAGTGGCCCCTTGATGTCGCACGCTT
       A   L   E   R   G   C   R   L   R   G   L   S   P   G   N   Y   S   V   R   I   -

2920  TCCCGGGCCACCTCCCTTGCGGGCAACGGCTCTTGGACGCGGAACCCACCTATTTCTACGTGA
      +---------+---------+---------+---------+---------+---------+  2979
      AGGCCCGGTGGAGGGAACGCCCGTTGCCGAGAACCTGCCTTGGGTGGATAAAGATGCACT
       R   A   T   S   L   A   G   N   G   S   W   T   E   P   T   Y   F   Y   V   T   -
                                                                    X
                                                                    b
                                                                    a
                                                                    I

2980  CAGACTATTTAGACGTCCCGTCAAATATTGCAAAATAATCTAGAGGATCTGGGGTGGCAT
      +---------+---------+---------+---------+---------+---------+  3039
      GTCTGATAAATCTGCAGGGCAGTTTATAACGTTTATTAGATCTCCTAGACCCCACCGTA
       D   Y   L   D   V   P   S   N   I   A   K   *   S   R   G   S   G   V   A   S   -
```

Fig. 4h

```
      CCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCCTGGAAGTTGCCCACTCCAGTGCCCACC
3040  +---------+---------+---------+---------+---------+---------+  3099
      GGGACACTGGGGAGGGTCACGGGAGGACCCGGACCTTCAACGGTGAGGTCACGGGTGG
       L  *  P  L  P  S  A  S  P  G  P  G  S  C  H  S  S  A  H  Q  -

      AGCCTTGTCCTAATAAATTAAGTTGCATCATTTGTCTGACTAGGTGTCCTTCTATAAT
3100  +---------+---------+---------+---------+---------+---------+  3159
      TCGGAACAGGATTATTTAATTCAACGTAGTAAAACAGACTGATCCACAGGAAGATATTA
       P  C  P  N  K  I  K  L  H  H  F  V  *  L  G  V  L  L  *  Y  -

      ATTATGGGTGGAGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAGACAACCTGTAGG
3160  +---------+---------+---------+---------+---------+---------+  3219
      TAATACCCCACCTCCCCCACCATACCTCGTTCCCCGTTCAACCCTTCTGTTGGACATCC
       Y  G  V  E  G  G  G  M  E  Q  G  A  S  W  E  D  N  L  *  G  -

      GCCTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGGCTCACTGCA
3220  +---------+---------+---------+---------+---------+---------+  3279
      CGGACGCCCCAGATAACCCTTGGTTCGACCTCACGTCACGTGTTAGAACCGAGTGACGT
       L  R  G  L  G  T  K  L  E  C  S  G  T  I  L  A  H  C  N  -

      ATCTCCGCCTCCTGGGTTCAA
3280  +---------+---------+  3300
      TAGAGGCGGAGGACCCAAGTT
       L  R  L  L  G  S  S  -
```

Fig. 4i
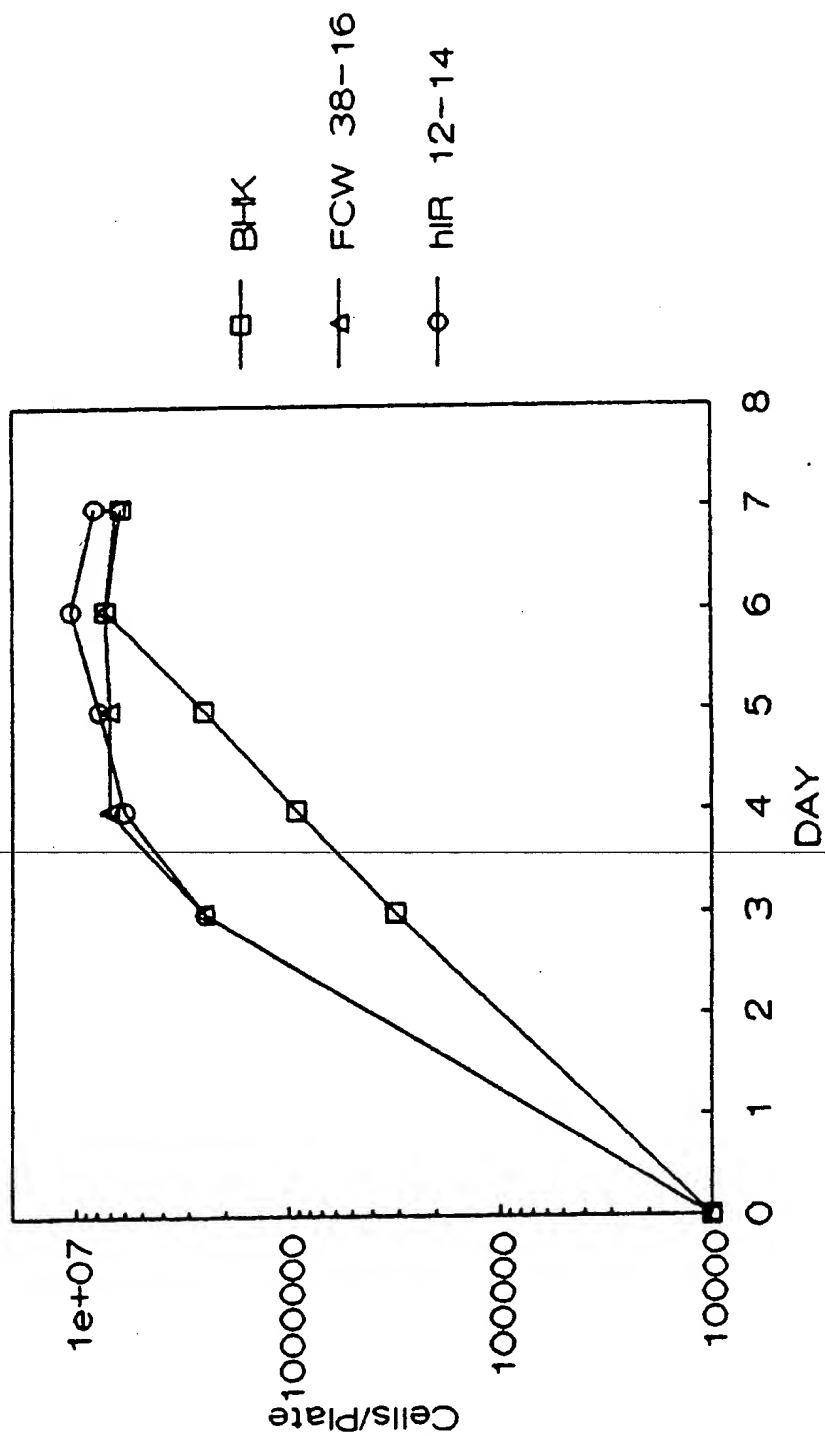
Fig. 5
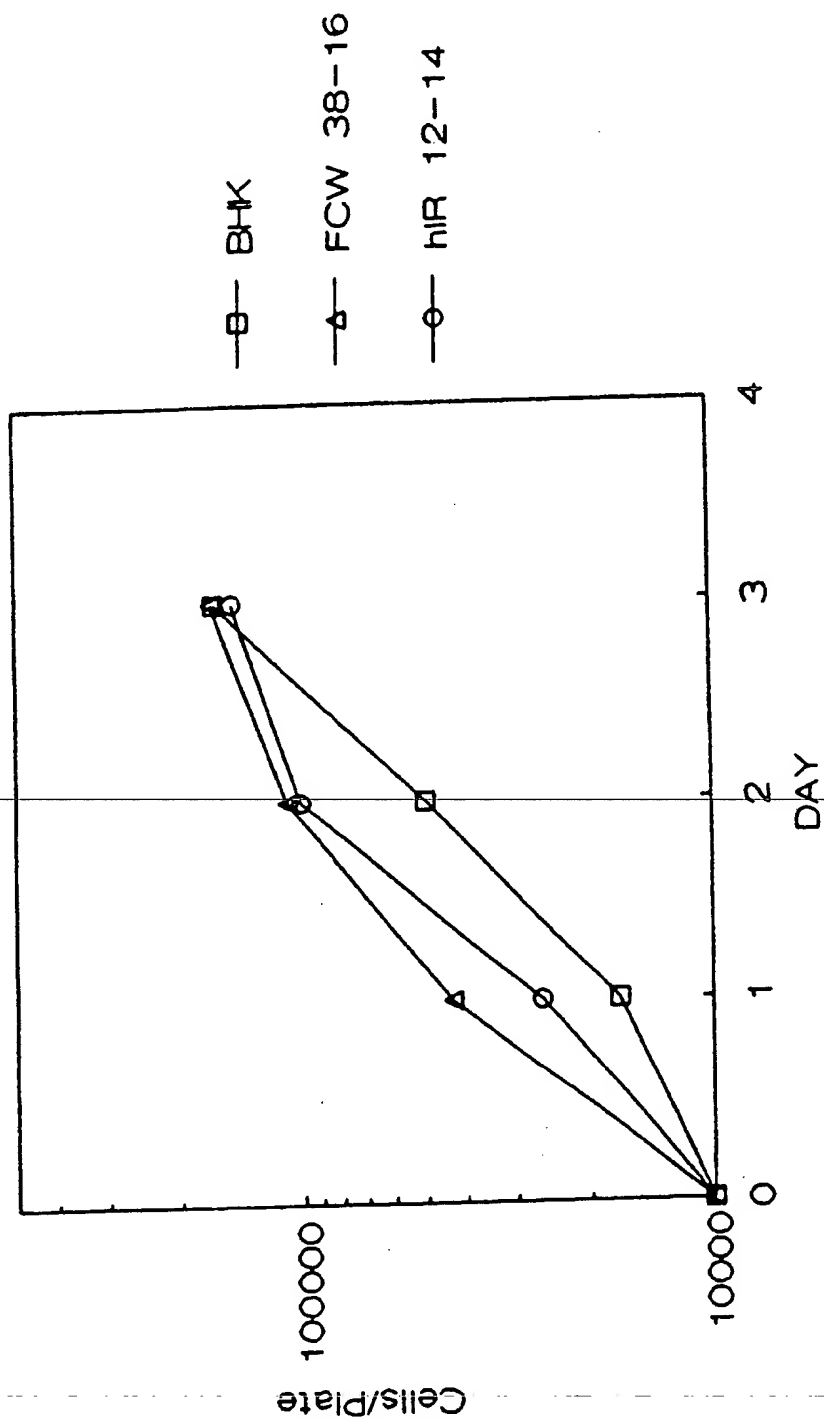
Fig. 6
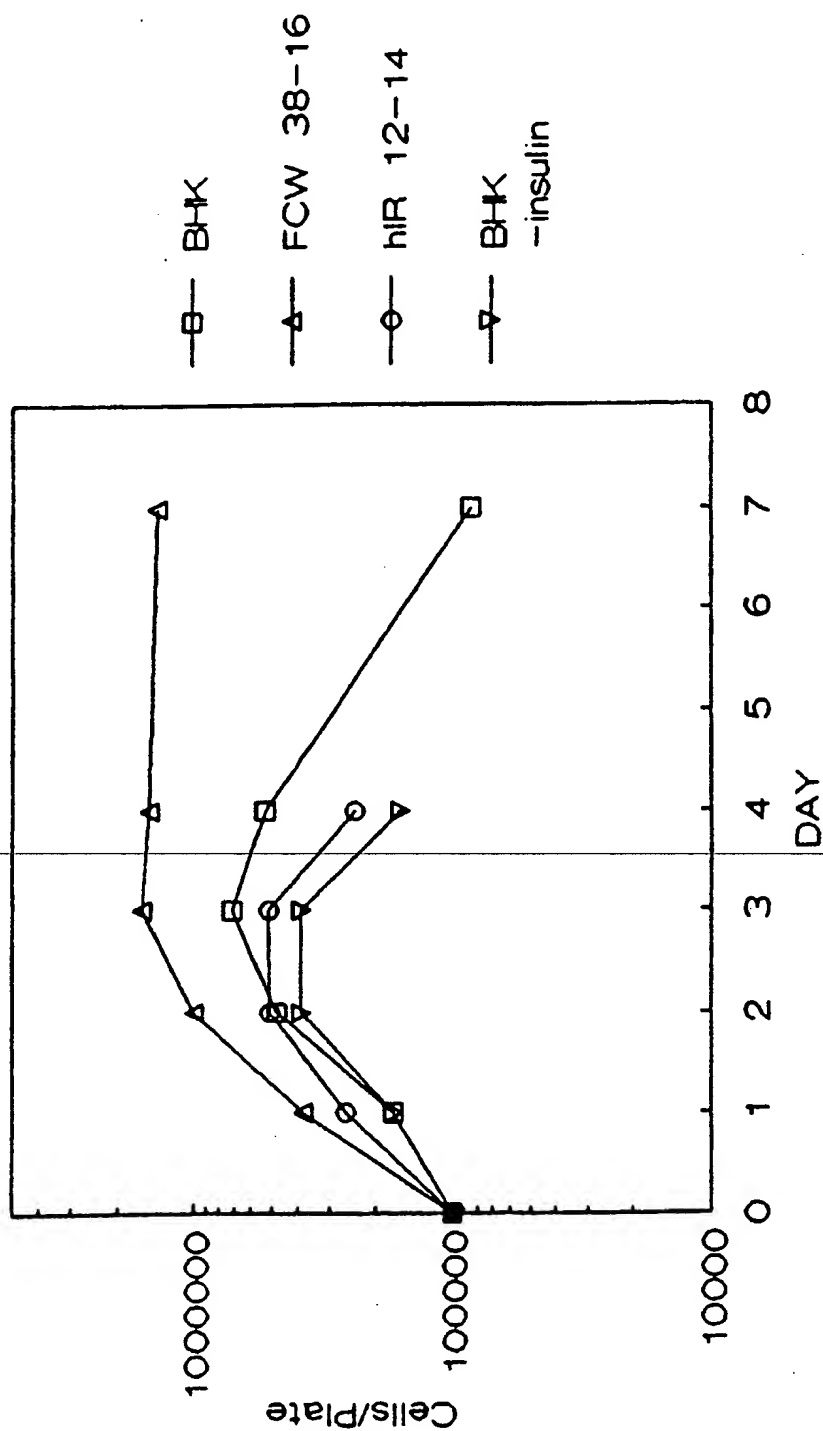
Fig. 7

INTERNATIONAL SEARCH REPORT

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC5: C 12 N 15/62, 5/00

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
IPC5	C 12 N; C 07 K; A 61 K

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in Fields Searched[8]

SE,DK,FI,NO classes as above

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category*	Citation of Document,[11] with indication, where appropriate, of the relevant passages[12]	Relevant to Claim No.[13]
Y	Proc. Natl. Acad. Sci. USA, Vol. 83, November 1986 L. Ellis et al.: "Linking functional domains of the human insulin receptor with the bacterial aspartate receptor", see page 8137 - page 8141	1-29
Y	WO, A1, 8905355 (THE SALK INSTITUTE FOR BIOLOGICAL STUDIES) 15 June 1989, see the whole document	1-29
Y	EP, A1, 0244221 (GENENTECH, INC.) 4 November 1987, see the whole document	1-29
Y	The Journal of Cell Biology, Vol. 109, November 1989 L. Sistonen et al.: "Activation of the neu Tyrosine Kinase Induces the fos/jun Transcription Factor Complex, the Glucose Transporter, and Ornithine Decarboxylase", see page 1911 - page 1919	28-29

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
17th July 1991	1991 -07- 2 5

International Searching Authority	Signature of Authorized Officer
SWEDISH PATENT OFFICE	Yvonne Siösteen

III. DOCUMENTS CONSIDERED TO BE RELEVANT   (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
A	Science, Vol. 240, June 1988 B.K. Kobilka et al.: "Chimeric α_2-, β_2-Adrenergic Receptors: Delineation of Domains Involved in Effector Coupling and Ligand Binding Specificity", see page 1310 - page 1316 --	1-29
A	EP, A2, 0325262 (THE GENERAL HOSPITAL CORPORATION) 26 July 1989, see the whole document --	1-29
A	Proc. Natl. Acad. Sci. USA, Vol. 84, August 1987 L. Ellis et al.: "Heterologous transmembrane signaling by a human insulin receptor-v-ros hybrid in Chinese hamster ovary cells", see page 5101 - page 5105 -- --------	1-29

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the Swedish Patent Office EDP file on   91-06-27
The Swedish Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A1- 8905355	89-06-15	AU-D-	2818889	89-07-05
		EP-A-	0325849	89-08-02
		US-A-	4981784	91-01-01
EP-A1- 0244221	87-11-04	JP-A-	62272990	87-11-27
		US-A-	4859609	89-08-22
EP-A2- 0325262	89-07-26	AU-D-	3281889	89-08-11
		WO-A-	89/06690	89-07-27